MUSEUM NATIONAL D'HISTOIRE NATURELLE

Ecole Doctorale Sciences de la Nature et de l'Homme – ED 227

Année 2013 **N°attribué par la bibliothèque**
|_|_|_|_|_|_|_|_|_|_|_|_|_|

THESE

Pour obtenir le grade de

DOCTEUR DU MUSEUM NATIONAL D'HISTOIRE NATURELLE

Spécialité : Biologie évolutive

Présentée et soutenue publiquement par

Yareli Sugey ESQUER GARRIGOS

Le 12 juin 2013

Multi-scale evolutionary analysis of a high altitude freshwater species flock: diversification of the *agassizii* complex (*Orestias*, Cyprinodontidae, Teleostei) across the Andean Altiplano.

Sous la direction de **: M. HUGUENY Bernard, Directeur de Recherche,**

et M. GAUBERT Philippe, Chargé de Recherche

JURY :

M. Philippe KEITH	Professeur, Muséum National d'Histoire Naturelle, Paris (075)	**Président**
M. Filip VOLCKAERT	Professeur, Katholieke Universiteit Leuven, Leuven (B)	**Rapporteur**
M. Jean-François AGNESE	Directeur de Recherche, Institut de Recherche pour le Développement, Montpellier (034)	**Rapporteur**
M. José A. GODOY	Senior Researcher, Estación Biológica de Doñana (CSIC), Sevilla (SP)	**Examinateur**
M. Marco MENDEZ	Associated Professor, Universidad de Chile, Santiago (CL)	**Examinateur**
M. Philippe GAUBERT	Chargé de Recherche, Institut de Recherche pour le Développement, Paris (075)	**Co-directeur de thèse**
M. Bernard HUGUENY	Directeur de Recherche, Institut de Recherche pour le Développement, Paris (075)	**Co-directeur de thèse**

Remerciements

Apres tous ces années, il faut faire l'effort et vous montrer mes habilites en français…qui sont pas du tout le top du top encore mais l'occasion vaut bien de faire l'effort. Donc, on y va…sans trucs, et pas 'goo… traducteur', juste moi et mon clavier.

Effectuer une thèse n'est pas une question simple. Il faut plusieurs ingrédients, dont : un esprit critique, de la curiosité, motivation, du temps (beaucoup!), de l'intégrité, le dévouement, et ... patience! Mais surtout des personnes, nombreuses personnes. Plusieurs personnes se sont impliqués au cours du ce travail, personnes sans lesquels ça m'aurait été pratiquement impossible de réaliser chaque étape, chaque objectif, quelque soit sa magnitude. De cette façon, je voudrais remercier à toutes ces personnes qui ont participé à l'évolution de mon travail, d'une manière directe ou indirecte, et aux qui je voudrais exprimer ma plus profonde et sincère gratitude. Un grand merci :

A mes directeurs de thèse, Philippe Gaubert et Bernard Hugueny, pour la confiance et le grand soutien qu'ils m'ont toujours donné, et plus encore pendant les moments plus difficiles…c'est à dire, la fin de la thèse ! Merci beaucoup a vous deux pour la qualité de votre soutien scientifique, pour toutes les discussions et les réflexions au cours de ces années. J'appris beaucoup de vous deux. Du même, votre qualité humaine et soutien moral au-delà des aspects spécifiques de la thèse c'est une chose que je n'oublierais jamais.

A Filip Volkaert et Jean-François Agnèse, rapporteurs, pour avoir accepté d'évaluer mon travail. Du même a Philip Keith, José A. Godoy et Marco A. Méndez, membres de mon jury de thèse, merci a tous.

Aux membres de mon comité de thèse non présents dans le jury, Thierry Oberdorff, Elie Poulin, Marie-Catherine Boisselier, Jawad Abdelkrim et Emmanuel Paradis pour les discussions et tous les conseils qui m'ont aidé a mieux orienté mon travail.

A Carla Ibañez pour tout son soutien depuis le début de ce travail et pour être toujours disponible pour répondre à toutes mes questions concernant les *Orestias* et l'Altiplano. Merci également à ses étudiants: Herbes Kelvin, Claudia Zepita et Alexander Flores pour les discussions et l'expérience d'avoir partagé avec eux quelques jours de travail sur le terrain. Ce n'est pas une question facile l'échantillonnage de ces poissons dans les conditions si particulières de l'Altiplano, mais vous le faites toujours avec une grande volonté et enthousiasme. Je voudrais aussi remercier Ramon Catari, qui a beaucoup soutenu le travail sur le terrain et de qui j'ai appris des anecdotes très intéressantes sur le lac Titicaca. Merci encore à tous, car votre travail constitue un élément très important de ce projet.

A Marco Méndez et Pamela Morales pour tout son soutien et disponibilité pour m'aider à améliorer la portée de cette recherche. Si j'ai la chance de continuer dans le domaine de la recherche, j'espère pouvoir retrouver des personnes si enthousiastes et disponibles pour travailler. Merci aussi pour la grande hospitalité pendant mon séjour dans votre beau pays. A Michel Sallabery et Hugo Salinas pour toutes les anecdotes et l'inoubliable expérience d'avoir partagé le terrain.

A Irma Vila, pour toute sa générosité, hospitalité et pour être toujours disponible pour discuter et partager toutes ses connaissances sur les *Orestias*.

A Laurent Lauzanne pour les enseignements et discussions utiles sur la taxonomie des *Orestias*, la validité de la morphologie et surtout sur le nombre d'espèces valides... question pour laquelle, même après presque quatre ans, je n'ai pas encore la réponse!

A Maldonado Mabel pour l'organisation de l'atelier 'Hacia la revisión sistemática del género *Orestias*' (Cochabamba, Bolivie) et qui a réuni plusieurs spécialistes de ce groupe emblématique de poissons et nous a permis de partager les différentes évidences et opinions concernant sa taxonomie et sa diversification. Merci de même a Michel Jegu, Soraya Barrera, Fernando Carvajal, et Evans de la Barra.

A l'Institut de Recherche pour le Développement, Département de Soutien et Formation pour la bourse de doctorat accordée, particulièrement à Sandrine Lerosier, Félicité Metonou y Nabil El Kente. Du même a EGIDE-Campus France pour la gestion de ma bourse et la sécurité social.

A l'Ecole Doctorale (ED 227 'Sciences de la Nature et de l'Homme') du Muséum National d'Histoire Naturelle, et particulièrement à Madiha Abada pour tout son soutien.

A l'UMR BOREA (Biologie des Organismes Aquatiques et Ecosystèmes), particulièrement à Sylvie Dufour et Thierry Oberdoff pour m'avoir accueilli dans cet unité et de même pour votre soutien. Aux membres de l'équipe 'Biodiversité et Macroécologie' pour votre soutien et tous les bons et nombreux moments partagés: Jean François, Clément Tisseuil, Murillo Sversut, Pablo Tedesco, Isabelle Mouas, Céline Jézéquel, Christian Lévêque, Rémy Bigorne et Didier Paugy. Merci a tous.

Au Service de Systématique Moléculaire du Muséum National d'Histoire Naturelle (Paris), particulièrement a Eric Pasquet pour m'avoir accueilli dans ce service. Du même à Annie Tillier, Josie Lambourdière, Céline Bonillo, José Utge pour tout leur encadrement technique, et en general a tous les membres de la famille 'SSM': Nicolas Puillandre, Nicolas Vidal, Sarah Samadi, Alexandre Hassanin, Carole Ferraira, Agnes Dettai, Isabelle Borges, Dario Zuccon, Delphine Grey, Julien Brisset. En espérant n'oublier personne, merci beaucoup a tous pour votre soutien, pour la amitié et autant de beaux moments partagés!

A Patrice Pruvost, Romain Causse pour nous avoir permis d'accéder à la collection des *Orestias* MNHN et pour tous leurs conseils. De même à Zora Gabsi pour aider à la gestion de la collection au cours des essais d'extraction de l'ADN et Claude Ferrara pour toutes les excellentes photographies de ces spécimens. Merci beaucoup à tous. Je voudrais aussi remercier à Régis Debruyne pour tous ses précieux conseils par rapport aux méthodes d'extraction de l'ADN ancien, mais aussi pour la gestion de la salle réservée à cet effet.

Aux 'les filles': Justine Thubaut, Anne-Claire Lautredou, Charlotte Schoelinck, Julie Marin avec qui j'ai partagé cette 'dur et terrible' marche mais avec qui j'ai partagé aussi des beaux moments de détente. Merci a toutes.

A mes très chers amis: Mariana Vargas, Sébastien Fromenteau, Sofie Vanmale, Sandra Rojo, Lee Galindo, Nicolas Nesi, Charles Muñoz, Stéphanie Leprieur, merci pour tous les moments partagés, les rires, les larmes ... eh voila, les larmes commencent à nouveau…enfin, merci pour tous les bons souvenirs.

A Jacqueline et Jean Cornu, merci pour votre soutien et pour m'avoir accueilli d'une manière si chaleureuse et affectueuse dans votre famille. Vous êtes et vous serez toujours très pressent dans mon coeur.

A mon 'JF'. Merci mon cœur pour tout ton soutien, pour être toujours là, pour écouter avec la plus grand attention toutes mes histoires de ces poissons, les discussions de la thèse, les plaintes, les questions…oh la la…mais c'est fini ! Ou pas…on verra ! En tout cas, la fin de cette étape et aussi bien grâce à toi, un vrai travail en équipe. Merci d'être toi, et d'être mon compagnon, tu le sais…

Et finalement, mais pas pour ça moins importants, je remercie ma mère, mon père et ma sœur, pour être toujours si présents, même dans la distance, avec tout cet amour et soutien inconditionnel. Je vous aime beaucoup.

Dédicace

A Jean-François Cornu et à la lumière de nos vies: Emilien.

Agradecimientos

La elaboración de una tesis no es asunto sencillo, se requiere de varios 'ingredientes' como: una gran motivación, tiempo (mucho!), entereza, dedicación, pasión, y…paciencia! Pero sobre todo de gente, mucha gente. Y aunque es uno mismo quien debe tomar el timón de su propio trabajo, una tesis no es solo asunto propio, muchas personas intervinieron a lo largo del desarrollo de este trabajo, personas sin las cuales me habría resultado prácticamente imposible el ir concretando con cada etapa, con cada objetivo, cualquiera que fuese su magnitud. Es así, que con estas líneas deseo dar las gracias a todas aquellas personas que participaron de forma directa o indirecta al desarrollo de mi trabajo, y a quienes expreso mi más sincera y profunda gratitud. Mil gracias:

A mis directores de tesis, Philippe Gaubert y Bernard Hugueny por la confianza y el gran apoyo que me brindaron siempre, y más aun durante los momentos mas difíciles de este trabajo…es decir, el fin de la tesis! Muchas gracias a ambos por la calidad de la supervisión científica, por las discusiones y reflexiones en el curso de estos años. He aprendido mucho de ambos. Así mismo, vuestra calidad humana y apoyo moral más allá de los aspectos propios de la tesis, es algo que nunca olvidaré.

A Filip Volkaert y Jean-François Agnèse, por haber aceptado evaluar mi tesis. Asi mismo a Philip Keith, José A. Godoy y Marco A. Méndez, miembros de mi jurado de tesis, gracias a todos.

A los miembros de mi comité de tesis, José Antonio Godoy, Thierry Oberdorff, Elie Poulin, Jawad Abdelkrim y Emmanuel Paradis por las discusiones tan productivas y que me ayudaron a centrar mejor mi trabajo.

A Carla Ibañez por todo el apoyo que me brindo desde el inicio de este trabajo y por estar siempre dispuesta a contestar mis preguntas sobre los *Orestias*. Gracias también a sus estudiantes: Kelvin Herbas, Claudia Zepita y Alexander Flores por las pláticas y la grata experiencia de compartir con ellos algunos días de trabajo en campo. No es asunto sencillo muestrear peces bajo las peculiares condiciones del Altiplano, pero ustedes lo hacen con la mayor disposición y entrega. De igual forma deseo dar las gracias a Ramón Catari, quien ha apoyado enormemente las colectas en campo y de quien aprendí interesantes anécdotas del lago Titicaca. Gracias nuevamente a todos, pues su trabajo forma una parte muy importante en el desarrollo de este proyecto.

A Marco Méndez y Pamela Morales por todo su apoyo y disponibilidad por ayudarme a mejorar el alcance de esta investigación. Si tengo la suerte de continuar por este camino, espero poder encontrarme con personas tan entusiastas y dispuestas por sacar los proyectos adelante. De igual forma les doy las gracias por la gran hospitalidad con la que me han recibido en su hermoso país. A Michel Sallabery y Hugo Salinas por las anécdotas y la grata experiencia de compartir el trabajo de campo.

A Irma Vila por su gran generosidad, hospitalidad y su gran disposición por compartir todo su saber sobre los *Orestias*.

A Laurent Lauzanne por las enseñanzas y valiosas discusiones sobre la taxonomia de los *Orestias*, la validez de los caracteres morfológicos y sobre todo acerca del número de especies…interrogante para la cual, aun después de casi 4 años, no tengo respuesta, tan solo apuestas!

A Mabel Maldonado por la organización del taller: 'Hacia la revisión sistemática del género *Orestias*' (Cochabamba, Bolivia) y que reunió a varios especialistas de este grupo emblemático de peces y permitió compartir y discutir las diferentes evidencias sobre la taxonomia y la diversificación. Igualmente doy las gracias a Michel Jegu, Soraya Barrera, Fernando Carvajal, y Evans de la Barra.

Al 'Institut de Recherche pour le Développement, Département de Soutien et Formation' por la beca de doctorado que me fue otorgada, especialmente a Sandrine Lerosier, Félicité Metonou y Nabil El Kente. Asi mismo a EGIDE-Campus France por la gestión de mi beca y seguridad social.

A la escuela doctoral (ED 227 'Sciences de la Nature et de l'Homme') del Museo Nacional de Historia Natural (Paris), especialmente a Madiha Abada por el apoyo brindado.

Al 'UMR BOREA (Biologie des Organismes Aquatiques et Ecosystèmes)', especialmente a Sylvie Dufour y Thierry Oberdoff por haberme aceptado en dicha unidad y el gran apoyo que me brindaron. A los miembros del equipo 'Biodiversité et Macroécologie' por su apoyo y todos los buenos momentos comartidos: Jean François Cornu, Clément Tisseuil, Murillo Sversut, Pablo Tedesco, Isabel Mouas, Céline Jézéquel, Christian Lévêque, Rémy Bigorne y Didier Paugy. Gracias a todos.

Al Servicio de Sistemática Molecular del Museo Nacional de Historia Natural (Paris), especialmente a Eric Pasquet por haberme recibido en dicho servicio. Así mismo a Annie Tillier, Josie Lambourdière, Céline Bonillo, José Utge por el gran apoyo técnico recibido, y en general a todos los otros miembros de la familia 'SSM': Nicolas Poulindre, Nicolas Vidal, Sarah Samadi, Alexandre Hassanin, Carole Ferraira, Isabel, Aignes Detai, Isabelle Borges, Dario Zuccon, Delphine Grey, Julien Brisset, espero no olvidarme de nadie. Gracias a todos por el apoyo, la amistad brindada y tantos gratos momentos compartidos!

A Patrice Pruvost y Roman Causse por permitirnos el acceso a la colección de los *Orestias* del MNHN, por todos sus consejos técnicos así como a Zora Gabsi por ayudarnos con el manejo de la colección durante las pruebas de extracción de ADN y a Claude Ferrara por las excelentes fotografías realizadas de estos especimenes. Muchas gracias a todos. Y ya que estamos en esta parte, aprovecho para dar las gracias a Regis Debruyne, por la asesoria brindada en técnicas de extracción de ADN antiguo en muestras problema y su gran labor en la organización de la sala reservada para este efecto.

A 'les filles': Justine Thubaut, Anne-Claire Lautrédou, Charlotte Schoelinck, Julie Marin con quienes compartí el andar de la tesis doctoral y con quienes disfrute de muchos momentos lindos. Gracias a todas.

A mis queridos amigos: Mariana Vargas, Sebastien Fromenteau, Sofie Vanmale, Sandra Rojo, Lee Galindo, Nicolas Nesi, Charles Muñoz, Stephanie Leprieur, gracias

por todos los momentos compartidos, las risas, los llantos…y allí lo tienen, las lágrimas comienzan a salir de nuevo…en fin, gracias por todos los bellos recuerdos.

A Jean y Jacqueline Cornu, gracias por todo su apoyo y por recibirme de una forma tan afectuosa en su familia. Siempre están y estarán muy presentes en mi corazón.

A mi ‘JF’. Gracias corazón por todo tu apoyo, por estar siempre a mi lado, por escuchar con la mayor atención todas mis historias acerca de estos peces, las discusiones de la tesis, las quejas, las preguntas…oh la, la…pero se terminó! o no…ya lo veremos. En todo caso, el fin de esta etapa la logré también gracias a ti, un verdadero trabajo en equipo. Gracias por ser tú, por ser mi compañero, tú lo sabes…

Y finalmente, pero no por ello menos importantes, quiero agradecer a mi madre, a mi padre y hermana, por estar siempre tan presentes, aun en la distancia, con el gran amor y apoyo incondicional que siempre me han dado. Los amo mucho.

Dedicatoria

A Jean-François Cornu y a la luz de nuestras vidas: Emilien.

Résumé étendu

Analyse multi-échelle d'un essaim d'espèces de haute altitude: diversification du complexe *agassizii* (*Orestias*, Cyprinodontidae, Teleostei) dans l'Altiplano andin.

I. Cadre et modèle d'étude.

Les essaims d'espèces —groupes monophylétiques d'espèces endémiques diversifiées— constituent des cas emblématiques pour l'étude de la diversification (e.g. les Cichlidae d'Afrique de l'Est). Deux principaux processus évolutifs à l'origine des essaims d'espèces ont été proposés, incluant la radiation adaptative (spéciation écologique rapide) et la radiation non-adaptative ('mutation-order speciation', plus lente).

Nous avons étudié un cas unique d'essaim d'espèces d'eau douce de haute altitude endémique du bassin inter-andin, à savoir le complexe d'espèces le plus diversifié au sein du genre *Orestias* : le complexe *agassizii*. L'aire de répartition principale du genre *Orestias* recouvre le lac Titicaca et le reste de l'Atiplano andin. En raison de sa grande diversité morphologique et écologique, le genre a traditionnellement été considéré comme un essaim d'espèces, dont les délimitations taxonomiques établies sur la base de critères morphologiques et méristiques ont toujours été débattues. Dans la dernière révision taxonomique incluant la série complète de spécimens types, 43 espèces et quatre complexes monophylétiques ont été delimités : *cuvieri* (4 espèces), *mulleri* (5 espèces) et *gilsoni* (10 espèces), endémiques du lac Titicaca et de lacs environnants, et *agassizii* (24 espèces) present dans le lac Titicaca mais aussi dans les autres systèmes hydrologiques du bassin inter-andin. Récemment, ces délimitations ont été remises en question par une analyse phylogénétique basée sur des séquences d'ADN qui ont montré que le complexe *agassizii* devait exclure le groupe *luteus* pour être phylogénétiquement valide. Par ailleurs, la monophylie réciproque de plusieurs espèces au sein du complexe *agassizii* n'a pas été retrouvée.

La connaissance des facteurs et mécanismes responsables de la diversification du genre *Orestias* est très fragmentaire. Certains auteurs ont suggéré que le genre a pu traverser

une phase de diversification importante au cours du Pléistocène, favorisée par un processus de radiation adaptative. La particularité du complexe *agassizii* que nous avons choisi d'étudier est que sa diversification, suivant un modèle de 'radiation', a pu avoir lieu dans deux contextes évolutifs distincts, à savoir (i) un contexte de diversification écologique avec compétition interspécifique et sans barrières aux flux de gènes à l'intérieur du lac Titicaca, et ii) un contexte de dispersion dans des habitats similaires en l'absence de compétition interspécifique à l'extérieur du lac. Ceci suggère que la radiation adaptative (ou spéciation écologique) n'est pas le seul processus qui pourrait expliquer la diversification du complexe, mais plutôt une combinaison de processus adaptatifs et non-adaptatifs.

II. Objectifs.

L'objectif général de mon travail a été d'élucider les mécanismes et les facteurs à l'origine de la diversification du genre *Orestias*, et plus particulièrement du complexe *agassizii*. En amont, nous avons d'abord réévalué les délimitations taxonomiques entre les complexes et les espèces d'*Orestias* en utilisant une approche phylogénétique moléculaire ; nous avons ainsi pu délimiter la position phylogénétique du complexe *agassizii* au moyen de caractères indépendants de la morphologie. Une deuxième étape de notre travail a été d'évaluer l'effet de l'hybridation sur la variation morphotypique au sein du complexe *agassizii* (à travers la caractérisation des hybrides *O. luteus* x *O. agassizii*), et de discuter le possible impact de ce phénomène sur les délimitations taxonomiques basées sur les critères morphologiques mais aussi sur la diversification morphologique du complexe. Finalement, nous avons évalué le rôle potentiel des fluctuations des paléolacs au cours du Pléistocène sur les processus de diversification du complexe *agassizii*, notamment en dehors du lac Titicaca, à travers d'un étude de génétique des populations à l'échelle de l'Altiplano andin. La synergie entre ces objectifs été consolidée par la mise en place d'une approche multi-échelle utilisant des séquences nucléotidiques (ADN mitochondrial et nucléaire), des marqueurs microsatellites, des critères morphologiques (morphométriques et méristiques) et des données

environnementales (altitude et distances hydrographiques), et faisant appel à des méthodes d'analyse couvrant les domaines de la phylogénie, de la phylogéographie et de la génétique des populations.

III. Matériel et méthodes.

Mon projet de thèse a bénéficié de deux collaborations internationales qui ont permis de soutenir l'effort d'échantillonnage (Universidad Mayor de San Andres, La Paz, Bolivia et Universidad de Chile, Santiago, Chili). La zone échantillonnée recouvre l'Altiplano au Pérou (zones satellites du Lac Titicaca), en Bolivie et au nord du Chili (15-23°S et 67-70°W), et représente plusieurs types d'habitats, tels que rivières, lacs, et 'bofedales' à une altitude située entre 3600 et 4800 m d'altitude. Au total, 2650 spécimens ont été prélevés parmi les représentants des quatre complexes traditionnellement délimités au sein du genre *Orestias*. Nous avons sélectionné un sous-échantillon de 1325 spécimens pour les analyses génétiques et morphométriques. La représentativité spécifique de nos analyses a été augmentée par l'inclusion de spécimens types du Muséum National d'Histoire Naturelle, notamment l'espèce la plus emblématique du genre, *O. cuvieri* (éteinte).

Une série de marqueurs moléculaires permettant d'adresser les différentes échelles d'investigation de notre thèse a été sélectionnée, incluant: la région de contrôle et le cytochrome *b* (ADN mitochondrial), le gène de la rhodopsine (ADN nucléaire) et 10 marqueurs microsatellites.

IV. Résultats et Discussion.

Au niveau phylogénétique, les analyses intégrant les spécimens types ont permis de délimiter plus avant le complexe *agassizii*, qui devrait exclure le groupe (ou complexe) *luteus* et *O. jussiei*. Le complexe *mulleri* est polyphylétique et le clade/complexe *gilsoni* est nommé de manière provisoire, sur la base de l'inclusion de morphotypes similaires aux espèces décrites au sein de ce complexe (aucun spécimen type n'a pu être inclus dans nos analyses). Eu égard

à la redélimitation des complexes par notre analyse phylogénétique, nos résultats remettent en question la validité des caractères morphologiques diagnostiques traditionnellement utilisés pour délimiter les complexes et les espèces d'*Orestias* et montre la nécessité urgente d'une réévaluation systématique intégrative au sein du genre. Par ailleurs, le patron phylogénétique 'deep and shalllow' observé au sein du genre (deux lignées divergentes capables de s'hybrider et se diversifiant en morpho-espèces non monophylétiques) correspond vraisemblablement à un scénario de spéciation écologique en cours au sein du Lac Titicaca commençant probablement au Pléistocène.

Nous avons montré l'existence d'hybridation naturelle entre *O. agassizii* et *O. luteus*, deux espèces présentes en sympatrie dans plusieurs lacs de l'Altiplano —dont le Titicaca. Les hybrides (F1) ne sont pas morphologiquement intermédiaires mais plus proches du morphotype parental *O. agassizii*. Cette proximité morphologique des hybrides suggère un effet positif de la sélection naturelle sur le morphotype apparemment 'plus performant' (en termes d'utilisation de l'habitat et de capacité de dispersion), ce qui semble être confirmé par les expériences de croisements en captivité. D'autre part, il est à noter que l'hybridation a pu jouer un rôle dans le 'chaos' taxonomique au sein du genre *Orestias*, notamment car les hybrides peuvent avoir un morpho-espace qui dépasse celui des espèces parentales.

Au niveau phylogéographique / génétique des populations, nos résultats suggèrent que les fluctuations des paléolacs et la connectivité entre les bassins ont eu un impact majeur sur la structure des populations au sein du complexe *agassizii*. Nous émettons l'hypothèse que les extensions maximales des paléolacs Ballivian à Tauca (0.6 Mya-14.5 kya) suivies d'évènements ultérieurs de dispersion vers des zones éloignées sont à l'origine de sa diversification à l'extérieur du Lac Titicaca, suivant une dynamique 'source-puits' influencée par l'isolement par la distance (notamment par les gradients altitudinaux). Le complexe est provisoirement défini comme un 'essaim d'espèces lacustro-riverin' résultant à la fois de processus de radiation adaptative au sein du lac Titicaca (i.e. spéciation écologique) et non-

adaptative dans l'Altiplano (i.e. 'mutation-order speciation'). De futures études éco-morphologiques et comportementales devraient aider à tester cette hypothèse.

En termes de conservation, la redéfinition phylogénétique des complexes d'*Orestias* a permis d'identifier *cuvieri* comme un clade vulnérable à l'extinction (*O. cuvieri* éteint) et composé de morphotypes spécialisés. Au sein du complexe *agassizii*, des 'evolutionary significant units' ont été délimitées dans des habitats fragiles de haute altitude au nord-est et au sud de l'Altiplano, mais les morpho-espèces endémiques chiliennes n'ont pas été confirmées. Le niveau précis de délimitation génétique atteint au sein du complexe *agassizii* et pour les hybrides (*O. agassizii x O. luteus*) devrait constituer un cadre utile à l'orientation des programmes de repeuplements des espèces d'*Orestias*.

Contents

Summarized list of figures

Summarized list of tables

Chapter I

- **Theoretical framework**
- **Study system**
- **Objectives**
- **Material and methods**

I.1 What is a freshwater species flock ?

A species flock refers to a monophyletic group of closely related, endemic species coexisting in a restricted geographic area, which in freshwater is generally a lake (Greenwood 1984; Salzburger & Meyer 2004). This term has been used since the beginning of the XXth century, first by botanists and was later adopted by zoologists studying the tremendous diversity of cichlid fishes in the African Rift lakes (Echelle & Kornfield 1984). Even if the definition of species flock may be subject to various interpretations, key features are common descent (i.e. monophyly, although not mandatory), coexistence in a same geographic area (i.e. endemism), a high number of cladogenic events over a relatively short period of time and a great level of eco-morphological diversification (Mayr 1984; Ribbink 1984). In freshwater systems, most of the species flocks (or species swarms) have been described from lacustrine environments (Seehausen 2004; Table 1), including both 'ancient' lakes —i.e. stemming from a continuous series of paleolakes; minimum age: approximately 120,000 years ago (Gorthner 1994; Martens 1997)— and more recent, postglacial lakes (Danley & Kocher 2001). Ancient and postglacial lakes are characterized by both steady limnological ageing processes and complex physical/chemical histories owing to drastic geological and climatic changes. These processes have played an important role in promoting diversification and speciation trough the opening of new ecological niches. Although major extinction waves from which resulted successive waves of colonization may also occur (Cristescu *et al.* 2010).

The study of lacustrine species flocks has tremendously increased our understanding of the mechanisms driving diversification and speciation at different evolutionary stages (i.e. from initial divergence to the completion of the speciation process). Fundamental theoretical developments regarding speciation, adaptive radiation

and hybridization have been —at least partly— based on the study of such systems (Martens 1997; Salzburger *et al.* 2002; Salzburger & Meyer 2004; Seehausen 2004; 2006; Cristescu *et al.* 2010; Von Rintelen *et al.* 2010). Considerable attention has been directed towards vertebrates, especially teleosteans; however numerous species flocks among invertebrates have been described (e.g. Amphipods, Ostracoda and Gastropoda, see Table 1). Within teleosteans, Cichlids (Perciformes: Cichlidae) from the East African lakes have been extensively studied and are considered prime examples of adaptive radiation and explosive speciation in lacustrine ecosystems (Schluter 2000; Kocher 2004; Seehausen 2006). In the latter, the number of endemic taxa may be enormous (>1000 species) (Salzburger & Meyer 2004). Time associated with diversification for many of these complexes are extremely short (i.e. <50,000 yrs). Despite such a short elapse of time, what characterizes this diversification is generally the large extent of trophic differentiation (Schluter 2000; Seehausen 2006), including benthic algae grazers, sand-dwelling planktivores, pelagic piscivores, egg predators and specialized pursuit predators. Although the "longevity" of ancient lakes has been seen as an important factor promoting high levels of biodiversity and endemism (Schön & Martens 2004; Albrecht & Wilke 2008), molecular evidence has suggested that the dynamics of diversification has been influenced by the rapid succession of fluctuations in lake level involving ecological instability (e.g. in the case of African cichlids, Johnson *et al.* 1996; Sturmbauer *et al.* 2001; Danley *et al.* 2012; Nagle & Simons 2012). Indeed, in lacustrine systems, the observed biodiversity patterns have been shaped by three fundamental processes —speciation, immigration and extinction (Schluter 2000; Schön & Martens 2004; Rundell & Price 2009)— that often took place within a short time frame compared to the age of the lakes.

Species flocks in marine environment seem to be much less common than in lakes or on islands (Duda and Rolán 2005) presumably because isolation of marine fauna is more difficult to achieve given the absence of physical barriers at fine geographical scales (Palumbi 1994), especially in groups with broad distribution and high dispersal abilities. Even so, some authors have proposed that the definition of species flock should not be restricted to lacustrine taxa. John and Avise (1998) applied the term to radiations of marine teleosteans for which evidence of rapid speciation had been provided. Examples include the Antartic perciform notothenioids and the North-eastern Pacific scorpaenids of the genus *Sebastes* (John and Avise 1998, Eastman and McCune 2000) and coral reef fishes in the Caribbean (McCartney *et al*. 2003). Within marine invertebrates, examples of species flock include the gastropod genus *Conus* from Cape Verde which occurs in a narrowly circumscribed geographical area unlike other marine species flocks (Duda and Rolán 2005). Sullivan *et al*. (2002), in delimiting a riverine species flock within west-central African mormyrids, used the term to circumscribe "a monophyletic assemblage of species, at least largely restricted to the geographical area of their origin (i.e., autochthonous), exhibiting a high level of sympatry, and rapid, or explosive, speciation relative to their nearest relatives in neighboring regions". In opposition to lacustrine species flocks, where dispersers are supposed to have invaded empty ecological niches (see §II.2), rivers are less likely to have provided environments free from teleostean communities for a group to diversify without interspecific competition. As a potential consequence, the rare cases of riverine species flocks that have been described to date are not characterized by a great eco-morphological diversity, such as in *Brienomyrus* (Sullivan *etal*. 2002).

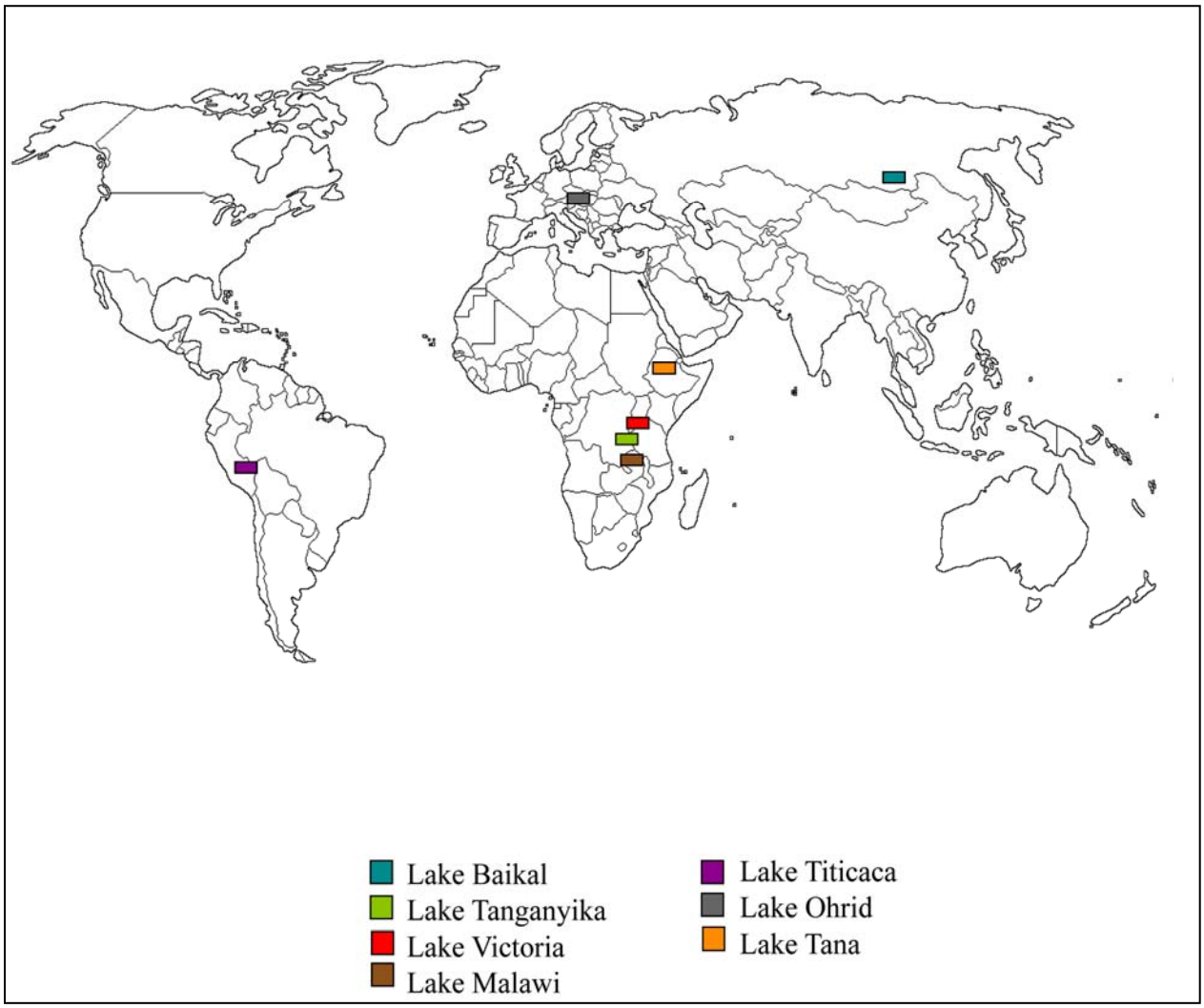


Figure 1. Geographic location of some of the main lakes where teleostean species flocks occur.

Table 1. List of some of the main lakes where teleostean species flocks occur. Data from: [(1)] Martens 1997 (1997), [(2)] Cristescu *et al.* (2010), [(3)] Kroll *et al.* (2012), [(4)] Albrecht & Wilke 2008 (2008) and [(5)] de Graaf *et al.* (2010).

Lake	Age (My)	Surface area (km^2)	Depth (max)	Teleostean species flocks	Invertebrates
Baikal (Russia)	25-70 (1,2)	31,500 (1,2)	1700 (1)	Cottoid (2)	Amphipods Ostracoda Isopoda Turbellaria Oligochaeta Gastropoda (2)
Tanganyika (East Africa)	5-20 (1,2)	32,600-32,893 (1,2)	1435 (1)	Cichlidae Bagridae Mastacembelidae Centropomidae Cichlidae Clupeidae Siluriformes (1,2)	Ostracods Harpacticoid Copepods Cyclopoids Potamonid Prosobranchs Decapoda Ostracoda Gastropoda Poriferans (1,2)
Malawi (East Africa)	2-5.5 (1,2)	29,604-30,380 (1,2)	785 (1)	Clariidae Cichlidae (1,2)	Prosobranchs Mollusca Ostracoda (1,2)
Victoria (East Africa)	0.75 (1,2)	68,635-70,000 (1,2)	93 (1)	Cichlidae (1,2)	-
Titicaca (South America)	2-3 (1,2,3)	8448 (1,2,3)	284 (1,2,3)	Cyprinodontidae (3)	Talitrid Prosobranchs Amphipoda Gastropoda Ostracoda Amphipods Microgastropods (1,2,3)
Lake Ohrid (Macedonia)	2-5 (1,4)	358 (1,4)	289 (4)	Salmonidae Cyprinidae Cobitidae Anguillidae (4)	Porifera Tricladida Plathelminthes Nemathelmintes Annelida Crustacea Mollusca (4)
Lake Tana (Ethiopia)	2-6 (5)	3050 (5)	15 (5)	Cyprinidae (5)	-

I.2 The processes behind species flocks: adaptive and non-adaptive radiations.

Mayr (1942) closely associated the concepts of species flock and explosive speciation (radiation). With the advent of molecular phylogenetics and later on, phylogeography, the complexity of evolutionary processes underlying species flock patterns were little by little unveiled. Adaptive radiation has been frequently identified as the main process leading to species flocks (Salzburger & Meyer 2004). The concept of adaptive radiation has been defined in two more or less generalizable ways: i) a pattern of species diversification in which the lineage of species occupies a diversity of ecological niches (e.g. Givnish 1997) and ii) the evolution of ecological and phenotypic diversity within a rapidly multiplying lineage. The classical view of adaptive radiation focuses on ecological opportunity, in which an ancestral species finds itself in an environment where resources are abundant and underutilized (Schluter 2000). Such resource availability may result from colonization of an under-populated area (e.g. island or lake), extinction of previously ecologically dominant groups, or evolution of a 'key innovation' (e.g. foraging traits in teleosteans, including particular head, jaw and gillraker configurations or shapes, that allow novel interactions with environment (Schluter 2000; Seehausen 2006; Streelman & Albertson 2006; Losos 2010).

Schluter (2000) proposed several criteria that would constitute a 'framework' to identify and/or validate cases of adaptive radiation within a clade of interest. According to the author, an adaptive radiation should be characterized by: i) rapid speciation rate, especially at early stage of diversification, ii) the adaptive nature of diversification, reflecting morphological, physiological and/or behavioral differences among species, and iii) ecophenotypes strongly correlated to biotic and abiotic characteristics of environment. However, there has been some debate over the definition of adaptive radiation, the main difference residing in the relative importance that authors assign to

speciation rate, species numbers and morphological diversity (see Sanderson 1998; Olson & Arroyo-Santos 2009). In such cases, the problematic characterization of adaptive radiation is likely due to the fact that defining the boundaries of an evolutionary process described as a continuous series of 'speciation stages' (see Fig.2; Nosil *et al.* 2009) is not straightforward. The term 'ecological speciation' has been proposed to define the completion of the process, i.e. speciation (reproductive isolation) via ecologically-based divergent selection (last stage on Fig. 2; Rundle & Nosil 2005; Rundell & Price 2009; Schluter 2009). Ecological speciation may be rapid, occurring across a few thousand years only (Hendry *et al.* 2007).

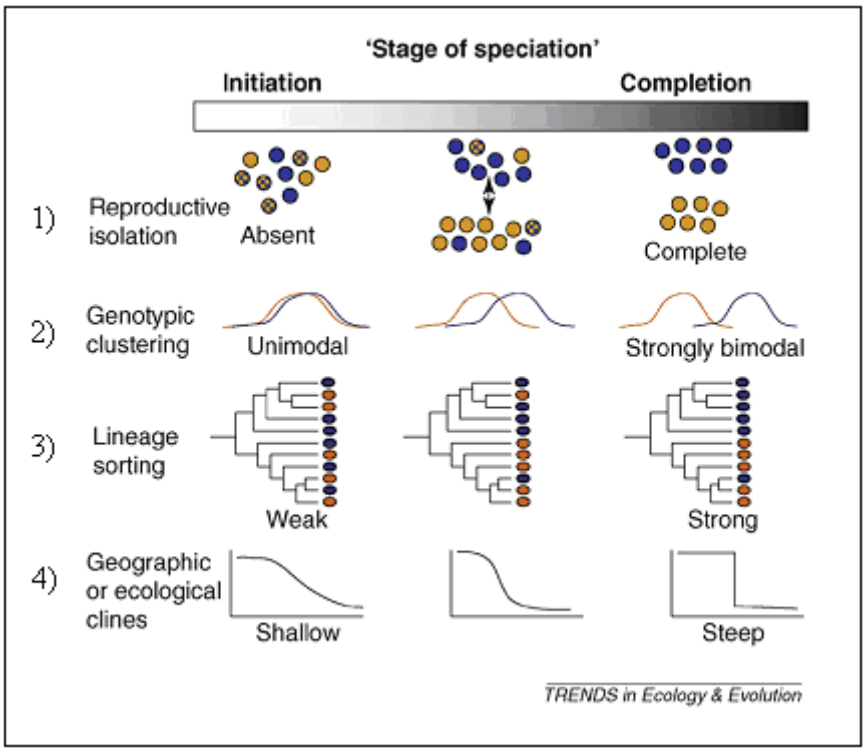


Figure 2. The continuous nature of divergence during speciation, from the initiation through to the completion of the speciation process (modified from Nosil *et al*. 2009).

Alternatively, non-adaptive radiation occurs when the rapid proliferation of species is accompanied by negligible or infrequent ecological differentiation or by morphological and physiological differentiation unrelated to patterns of resource use and environment (Gittenberger 1991; Rundell & Price 2009). Non-adaptive radiation is defined as evolutionary diversification "from a single ancestor with minimal ecological diversification, resulting in a series of ecologically-related similar species that are allopatric or parapatric replacements of one another", with implicit reference to deep evolutionary divergence (several millions of years; Rundell & Price 2009). Indeed, several case studies have suggested that phylogenetic niche conservatism contributes to the rapid accumulation of lineages through vicariant isolation and spatio-temporally fluctuating environments (Kozak *et al.* 2006; Nagle & Simons 2012), sometimes leading to eco-morphologically undifferentiated species (Sullivan *et al.* 2002). Such process where genetic drift acts in an allopatric context can be related to the "mutation-order speciation", which consists in the fixation of mutations in separate populations adapting to similar selection pressures, as opposed to ecological speciation (Schluter 2009).

Interestingly, several cases of radiation seem to be the product of both non-adaptive and adaptive processes, since they both contain ecologically differentiated, sympatric species and ecologically similar, allopatric or parapatric species (Gillespie 2004). Moreover, Rundell & Price (2009) predicted that species originating from ancient, non-adaptive radiation can later follow a rapid ecological differentiation when a change in environmental conditions make new, diversified niches available, thus producing the pattern of an adaptive radiation. In both the framework of adaptive (1) and non-adaptive radiations (2), hybridization among species may regularly occur during the invasion of new habitats or when secondary contact zones among species

occur, respectively. In (1), hybridization may contribute to increased rates of response to selection, thus predisposing new colonizers to rapid adaptive diversification under disruptive or divergent selection (hybrid speciation; Seehausen 2004). In (2), the consequences of hybridization are poorly known, but may also play a role in the diversification process by rearranging the genomic landscape (altough a deleterious role has been envisaged; Kozak *et al.* 2006).

I.3 The study system – part I: the genus *Orestias* (Cyprinodontiformes, Teleostei).

I.3.1 Evolutionary origin of *Orestias*

The genus *Orestias* belongs to the Order Cyprinodontiformes, commonly referred as killifishes. Cyprinodontiformes comprise ray-finned and surface-dwelling species that represent the most speciose assemblage of atherinomorph teleosteans, with over 1013 —mostly— freshwater species and ten families distributed circumtropically (west of Wallace's Line) and in some temperate regions of North America and Europe (Fig. 3). Some species of Cyprinodontiformes also have the ability to reside in brackish environments. They are found in a wide variety of habitats from isolated springs to lakes, rivers and marine waters. Members of this order comprise mostly small (ca. 0.8 to 5.0 cm) species, although larger species also exist, such as *Anableps microlepis* (Anablepidae) with ca. 34 cm of total length (Berra 2001; Nelson 2006).

Cyprinodontidae, commonly referred as pupfishes, are the largest and most diverse family within Cyprinodontiformes, with nine genera and over 100 described species (Fig. 4). Members of this family are distributed in North, Middle, and South America, the Caribbean, the Mediterranean region (including North Africa, Spain, Italy, Turkey, Greece and Mediterranean islands), as well as the Arabian Peninsula and Iran (Fig. 3). They mainly inhabit freshwater systems but also salt marshes, mangroves,

estuaries and coastal marine environments. Members of this family are oviparous. Three genera are remarkably speciose: *Cyprinodon*, *Aphanius* and *Orestias* (Parenti 1981; Nelson 2006). Within this family, the tribe Orestiini is composed of three genera, including: i) *Orestias* (47 species) distributed in high-altitude lakes of the South American inter-Andean basin, ii) *Aphanius* (10 species) distributed in the Mediterranean region, the Arabian Peninsula and Iran, and iii) *Kosswigighthys* (4 species), distributed in Turkey. This tribe represents one of the most notable cases of disjunct distribution in postulated monophyletic freshwater taxa (Parenti 1981; Parker & Kornfield 1995; Costa 1997; Berra 2001; Nelson 2006). It was defined by Parenti (1981) as monophyletic on the basis of a single synapomorphic character: a broadly expanded medial process of the dentary, giving a characteristically robust lower jaw.

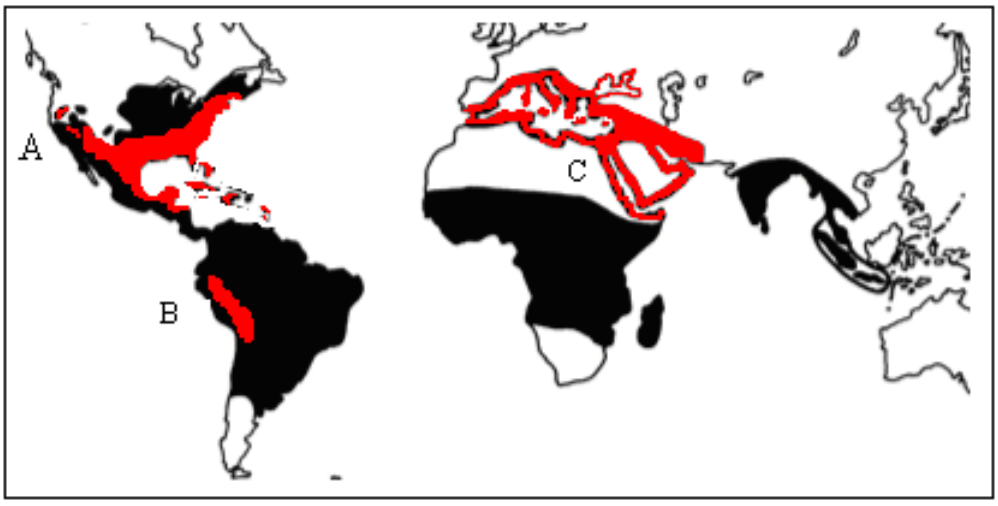


Figure 3. Geographic distribution of Cyprinodontiformes (in black), including Cyprinodontidae (in red): A) North America and the Caribbean Islands (*Cubanichthys*, *Cyprinodon*, *Jordanella*, *Cualac*, *Floridichthys* and *Megupsilon*), B) Inter-Andean basin (*Orestias*), and C) Mediterranean and Arabian Peninsula borders (*Aphanius* and *Kosswigichthys*) (modified from Parenti 1981; Parker &Kornfield 1995).

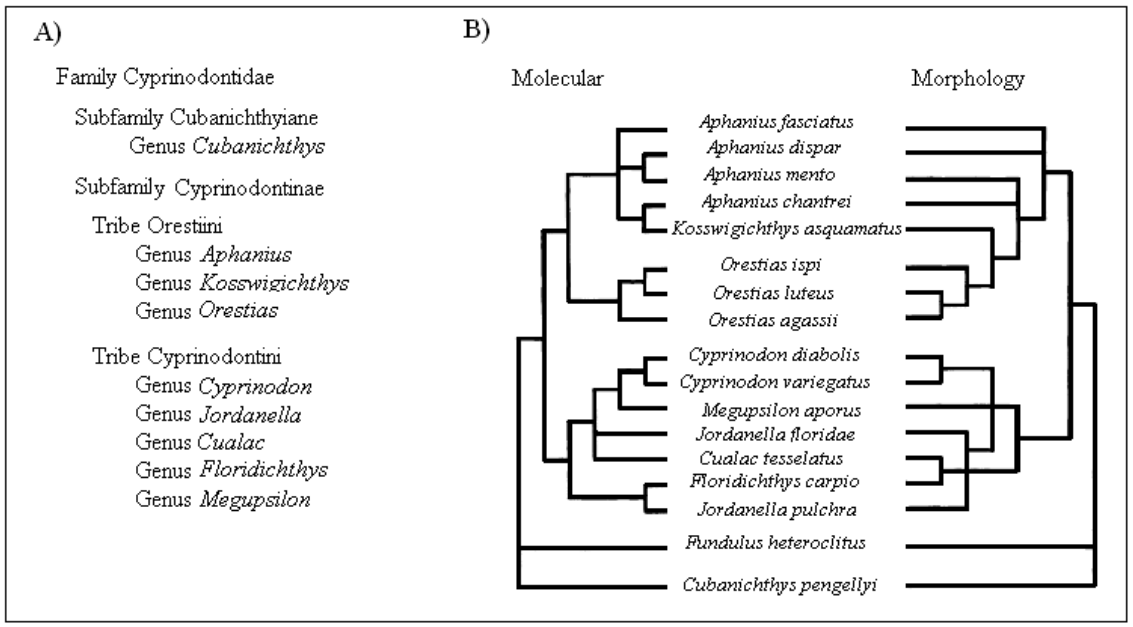


Figure 4. Family Cyprinodontidae: A) genus-level classification after Parenti (1981), B) molecular (Parker & Kornfield 1995) and morphological phylogenies (Parenti 1981) (modified from Parker & Kornfield 1995).

The phylogenetic relationships of the genus *Orestias* have been the subject of many discussions. Garman (1895) and Eigenmann (1910) hypothesized an origin from a fundulid-like ancestor. Later, Eigenmann (1920) considered this genus to be closely related to *Empetrichthys* (Goodeidae) on the basis of the absence of pelvic fins, whereas Foster (1967) classified it as closely related to *Cynolebias* (Rivulidae) on the basis of similarities in the supraorbital neuromast system. Alternatively, some authors placed the genus on its own (Regan 1911; Myers 1931). Parenti (1981) hypothesized that the genus *Aphanius* and *Kosswigichthys* were the sister group of genus *Orestias*, based on the observation of osteological characters. This morphological hypothesis was later confirmed by the phylogenetic analysis of mitochondrial DNA sequences (Parker & Kornfield 1995) (Fig. 4).

The inferred sister group relationship among the three genera has prevented any convincing explanation about their present-day disjunctive biogeographic pattern. The observed pattern could be derived from high levels of extinction (Parenti 1981) or events of great dispersion (Villwock & Sienknecht 1993). Parker and Kornfield (1995) posited that ancestral cyprinodontids inhabited the western Tethys Ocean in the Late Jurassic to Early Cretaceous (Fig. 5). As the North Atlantic basin filled, new habitats became available for colonization. At that time some of these ancestral cyprinodontids would have had a first opportunity to disperse to North America and the Caribbean, where they were later isolated from their Tethyan relatives about 145 to 66 Mya. The ocean sea level reached its maximum 100 to 88 Mya and permitted the formation of epicontinental seas that covered a large land area of Africa, North America and northern South America. The westernmost extent of this new region was the slopes of the eastern Andean Cordillera, which later began a period of uplift that reached its maximum activity about 25 Mya (Decou *et al.* 2013). At this time, Tethyan cyprinodontids had their second dispersal opportunity. Some of these cyprinodontids occupied drainages peripheral to their coastal habitats; following drainage shifting and stream capture during periods of active orogeny about 15 Mya (Van Der Voo 1993), they could have been able to disperse southward to the Lake Titicaca basin. However, the scenario proposed by these authors might prove problematic, since it posits the existence of ancestral cyprinodontids prior to the appearance of the atherinomorph fishes in the fossil record (Upper Cretaceous; see Parker & Kornfield 1995). Regarding *Orestias*, no fossil record exists. However, a recent re-evaluation of the fossil species *Carrionellus diumortuus* from the Lower Miocene of Ecuador supported its inclusion within the tribe Orestiini as the closest fossil representatives to *Orestias* (Costa 2012).

Figure 5. Habitat continuity and range of putative ancestors of *Orestias* and Anatolian cyprinodontiforms (Africa and South America), at the beginning of the Cretaceous. Areas covered by epicontinental seas during this period are in gray (modified from Parker & Kornfield 1995). North American cyprinodontiforms represent putative ancestors of the non-Orestinii lineage, isolated by vicariance with the opening of the North Atlantic basin.

I.3.2 Taxonomy of *Orestias*

The genus *Orestias*, so-called Andean pupfishes, has been defined as a monophyletic group on the basis of seven morphological synapomorphies (Parenti 1984b): (i) pelvic fins and fin girdle absent, (ii) vomer absent, (iii) middle anal and middle dorsal fin radials cartilaginous rather than ossified, (iv) anterior and posterior ceratohyal separated ventrally by a large gap filled with cartilage, (v) first postcleithrum absent, (vi) anguloarticular without ventral extension parallel to the retroarticular, and (vii) squamation and head pore pattern characterized by a prominent lyre-shaped arrangement of minute neuromasts and a prominent median dorsal ridge of scales running from the top of the head to the dorsal fin origin.

Taxonomic delimitations within *Orestias* by means of classical taxonomic characters (i.e. morphometrics and meristic counts) have long been chaotic (Villwock 1986; Loubens 1989; Lauzanne 1992; Müller 1993; Villwock & Sienknecht 1996; Lüssen *et al.* 2003). The first large collection of *Orestias* was brought to France by Pentland in 1839 and was deposited at the Muséum National d'Histoire Naturelle of Paris (Tchernavin 1944). From this collection, Valenciennes (1839) described the genus *Orestias* and ten species (in Cuvier & Valenciennes 1846). Since then, and following the analysis of a second large collection (Agassizz and Garman, deposited in the Museum of Comparative Zoology of Harvard and analized latter by Garman 1876) the number of species delimited within the genus varied significantly (see Article 1 for more details), from five to 20 (Günther 1866; Garman 1876;1895; Tchernavin 1944). In the last and most exhaustive taxonomic revision based on the analysis of morphometric, osteological and meristic characters, including the complete series of type material, Parenti (1984b) listed 43 species and delimited four monophyletic complexes (Fig. 6): *cuvieri* (4 species), *mulleri* (5 species) and *gilsoni* (10 species), endemic to Lake Titicaca and nearby water bodies, and the "*agassii*" complex (24 species) found in Lake Titicaca but also in other hydrological systems within the inter-Andean basin. There has been some level of confusion on the right spelling of the specific epithet to use for the name-bearing species of the latter complex (*agassii*, *agassi*, *agassisii* or *agassizii*). Following Eschmeyer & Fong (2013). I will use here the spelling *agassizii* because (i) in the original description of Valenciennes (Cuvier & Valenciennes 1846), the species is first referred as "*Orestias agassizii"* and (ii) the first reviser of the genus (Garman 1895) selected the epithet *agassizii*. Three additional species from Chile and belonging to the *agassizii* complex have been described since then (Vila & Pinto 1986; Vila 2006; Vila *et al.* 2011), accounting for a total of 27 species, from which 7 are endemic from

Lake Titicaca, 17 are outside the lake and three of them are found both in and outside Lake Titicaca (Fig. 6).

Synaphomorphic characters supporting the monophyly of each complex were described by Parenti (1984b). The *mulleri* complex was defined by a single synaphomorphy, namely the base of the dorsal and the anal fin projecting beyond the primarily body profile and being usually covered with small and thin scales. The *gilsoni* complex was defined also by a single synapomorphy, namely the caudal fin rays lying mostly interior to the body profile and closer to the vertebrate column, giving the caudal fin a 'tapered' external appearance. The *agassizii* complex was defined by three synaphomorphies, namely lateral scales from the operculum to the posterior extent of the pectoral fin enlarged with concentric striae absent or present to a limited degree, caudal peduncle relatively deep in adults, and scales on the dorsal portion of the head enlarged and with no or few discernible concentric striae. Within the *agassizii* complex a subgroup of five species (the '*luteus*' group) was delineated sharing a derived squamation pattern (i.e. thick and granulated scales on the head). The *cuvieri* complex included four species and was defined by seven synapomorphic characters (see Parenti 1984b, p. 153). According to Parenti (1984b), the *mulleri* and *gilsoni* complexes are sister groups, closely related to the *cuvieri* complex. These three complexes together form the sister group of the *agassizii* complex.

Recently, the morphological-based systematics of the genus was partially revised and severely questioned through a phylogenetic analysis using mitochondrial DNA sequences (Lüssen *et al.* 2003). The authors failed to recover the monophyly of the *agassizii* complex as proposed by Parenti (1984b), and showed that it should exclude *O. luteus* to be phylogenetically valid. Besides, the reciprocal monophyly of several species within the *agassizii* complex was not recovered.

The long-standing difficulty to discriminate among *Orestias* species may be due to several factors (Table 2), here defined as i) extrinsic (i.e. factors inherent to taxonomic practice or 'human-induced'), and ii) intrinsic (i.e. biological characteristics inherent to *Orestias*). In addition, several cases of hybridization among species/complexes of *Orestias* have been reported: *O. olivaceus* x *O. agassizii* (*O. affinis*), *O. cuvieri* x *O. pentlandii* and *O. agassizii* x *O. luteus* (Tchernavin 1944; Villwock & Sienknecht 1995), complicating further the delineation among morpho-species (see Article 3). In the case of *Orestias*, a remarkable combination of "unfavorable" factors including: taxonomic practices, permeable species boundaries due to probable radiation-like evolutionary processes and/or ecological plasticity, has long hampered the establishment of clear-cut morpho-species boundaries.

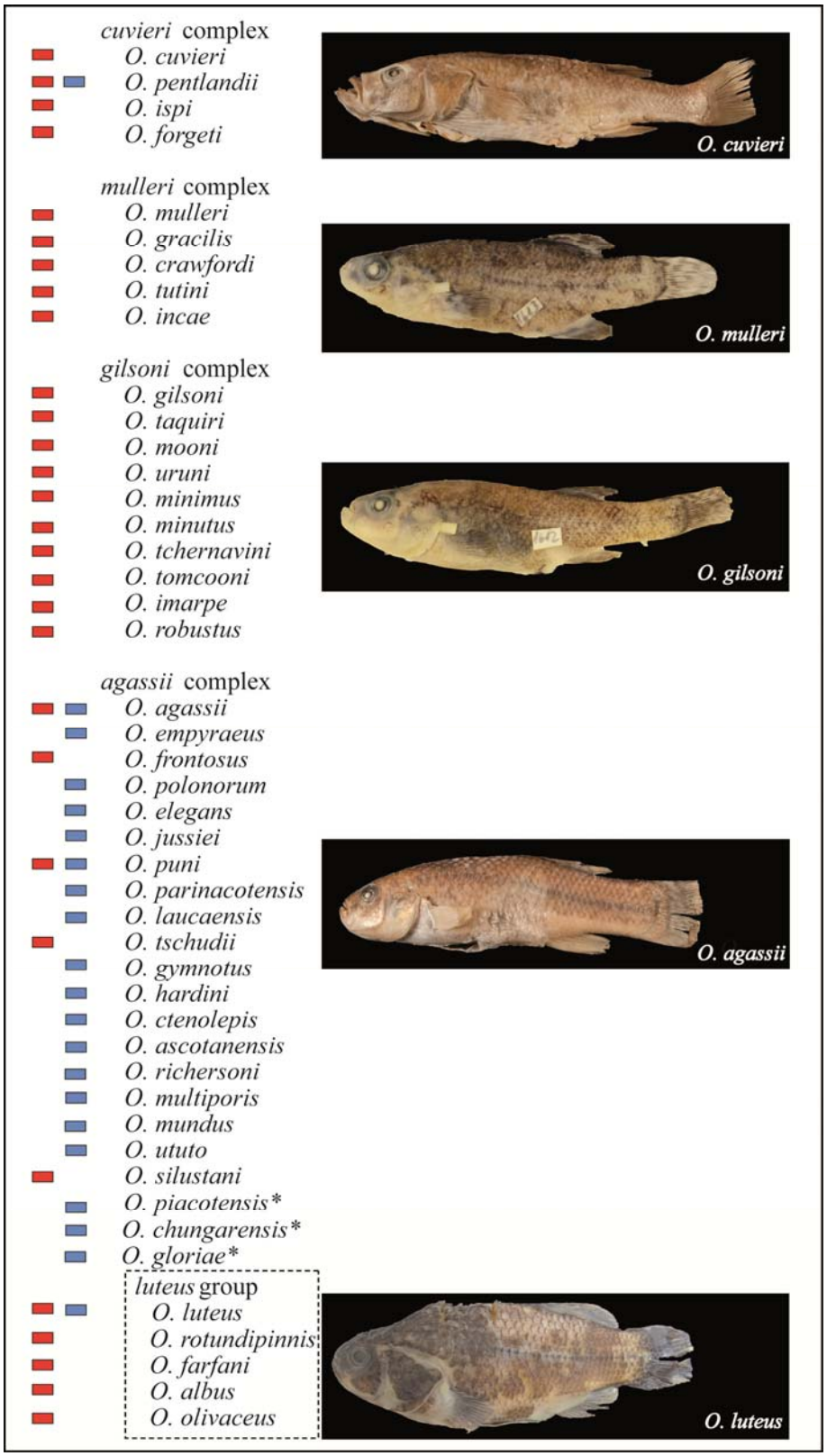


Figure 6. Complexes and species as proposed by Parenti (1984b). Red squares represent species endemic to Lake Titicaca, and blue squares represent species occurring outside Lake Titicaca. * Chilean species of the *agassizii* complex, described after Parenti (see Vila & Pinto 1986; Vila 2006; Vila *et al*. 2011).

Table 2. Extrinsic and intrinsic factors hindering the taxonomic delimitation of *Orestias* species.

Extrinsic factors:	Potential intrinsic factors:
a) Species description based on a single or a few specimens (e.g. *O. minutus*, described only from two specimens). b) Species description based on juvenile series or body size groups (e.g. *O. olivaceus* possibly representing the juvenile stage of *O. luteus*). c) Species description based on poorly preserved specimens (e.g. *O. tutuni*, known only from a warped and poorly preserved type series). d) Lack of uniformity on terms used for the descriptions (e.g. "the lateral line used by Garman is not the lateral line used by Cope"; in Tchernavin 1944, p. 146). e) No detailed analysis of intra- and inter-specific morphological variation at a large geographic scale, especially within the *agassizii* complex. f) Absence of currently collectable "species": extinct species (*O. cuvieri*) or described species not found or identifiable in later surveys (e.g. *O. tutini*, *O. taquiri*, *O. pentlandii*, *O. forgeti*).	a) Recent origin of the genus → no effective reproductive barriers between morpho-species b) Incidence of hybridization → intermediate forms (e.g. *O. cuvieri* x *O. pentlandii*) c) Sexual maturity at early developmental stage (small body size) → species described on the basis of juvenile proportions, scaling and colouring. d) Characters varying with age and growth (e.g. scaling, colouring) → species described from 'unstable' morphotypic series on an ontogenetic point of view. e) Ecological plasticity → a same species may show several ecotypes, depending on the type of habitats.
→ Reliable identification keys not available	

I.3.3 Distribution and ecological characteristics of *Orestias*

The genus *Orestias* is endemic to the high-altitude hydrological systems of the inter-Andean basin, and is distributed from the Ancash Province in northern Peru to Antofagasta Province in northern Chile (Parenti 1984b; Vila *et al.* 2007b). The main area of its distribution covers the Altiplano, including the Lake Titicaca basin on the border between Peru and Bolivia (Fig. 7), containing 60 % of the described species, of which 90 % are endemic to the lake (Lauzanne 1982; Parenti 1984b; Villwock 1986; Lauzanne 1992; Vila *et al.* 2007b). However, the number of species is almost equally represented outside the lake (Fig. 6), mostly with allopatric distributions (*versus*

sympatric distribution within the lake). The genus *Orestias* represents one of the three native teleostean genera present in the inter-Andean basin, along with the catfishes *Trychomicterus* and *Astroblepus* (Parenti 1984b; Lauzanne 1992; Vila *et al.* 2007a). It might be considered as an ecologically successful genus, not only regarding its great number of species (Fig. 6) but also because of its occurrence in a wide variety of habitats (Fig. 8), including small and isolated springs draining into salt pans ('salares'), lakes, rivers, wetlands ('bofedales') and small lakes at altitude levels varying between 3500 and 5200 m (Parenti 1984b; Vila *et al.* 2007b).

Orestias has diversified into a variety of morpho-species (Fig. 9), including inshore forms with deep bodies and slender offshore forms (Parenti 1984, Lauzanne 1982, Maldonado *et al* 2009, NorthCote 2000). Data on the trophic ecology of *Orestias* are scarce and only concern species from Lake Titicaca having some importance for the commercial fisheries (see Fig. 9). Some of these species show a segregation pattern in niche occupancy (Fig. 9 and Table 3), some feeding heavily on zooplankton (e.g. *O. pentlandii, O. ispi, O. mulleri*), others feeding preferentially on molluscs (*O. luteus*) and amphipods (*O. albus*), while others feed on Cladocera and algae (*O. agassizii*). These species present a wide range in trophogastric morphology (e.g. gill raker number, spacing and length of gill raker, pharyngeal dentition, alimentary canal length) that seems to be correlated to differences in feeding strategies (Lauzanne 1982; Lauzanne 1992; Northcote 2000; Maldonado *et al.* 2009). Nevertheless, it appears that competition for resources among *Orestias* morpho-species does occur within Lake Titicaca, especially in the littoral zone (Lauzanne 1982; Loubens 1989; Lauzanne 1992; Northcote 2000). *Orestias agassizii* has the broadest spectrum of food categories and is considered a generalist species (Lauzanne 1992).

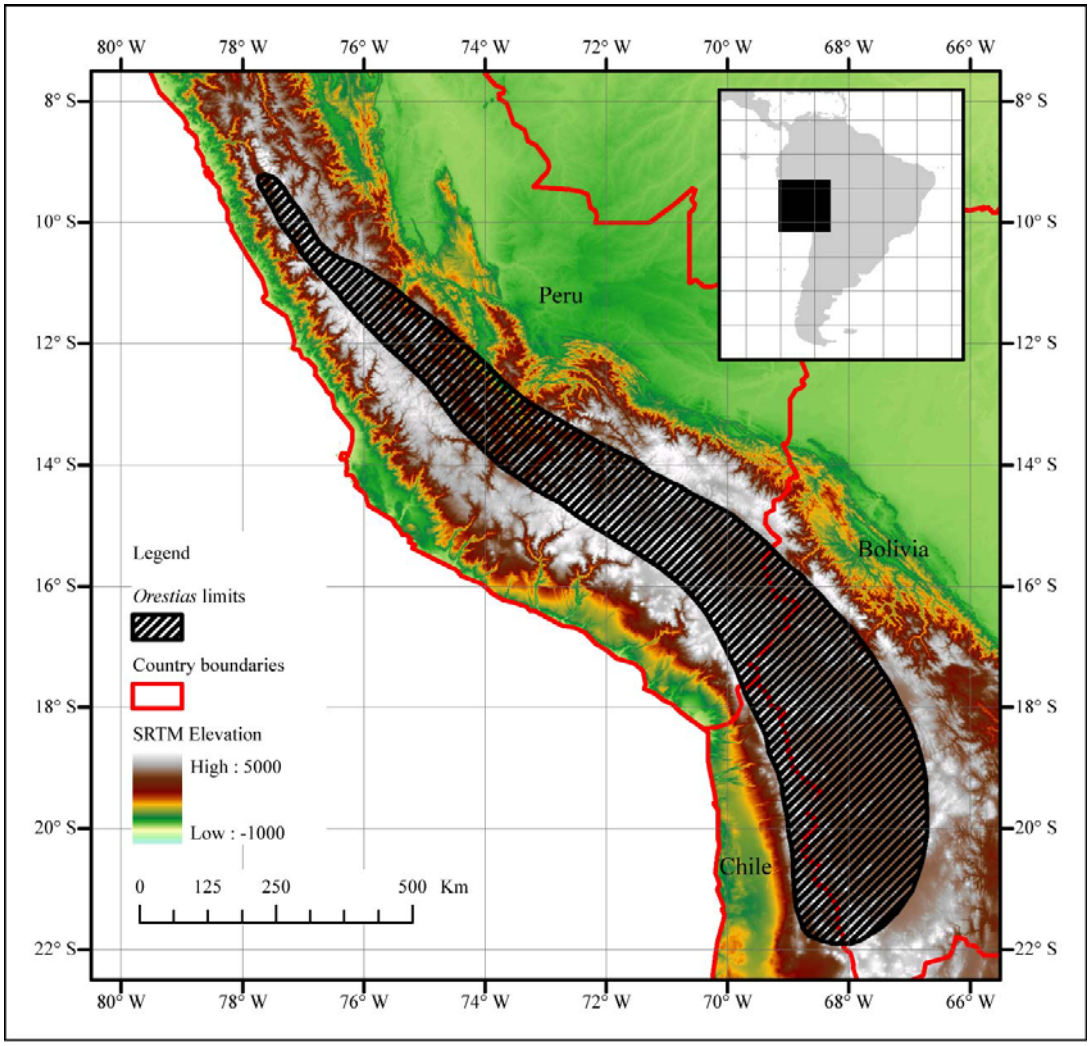


Figure 7. Geographic range of the genus *Orestias* (cross-hatched area).

Figure 8. Types of habitats for *Orestias*: A) spring immersed within salt pans ('salares'), B) lake, C) river, D) wetlands ('bofedales').

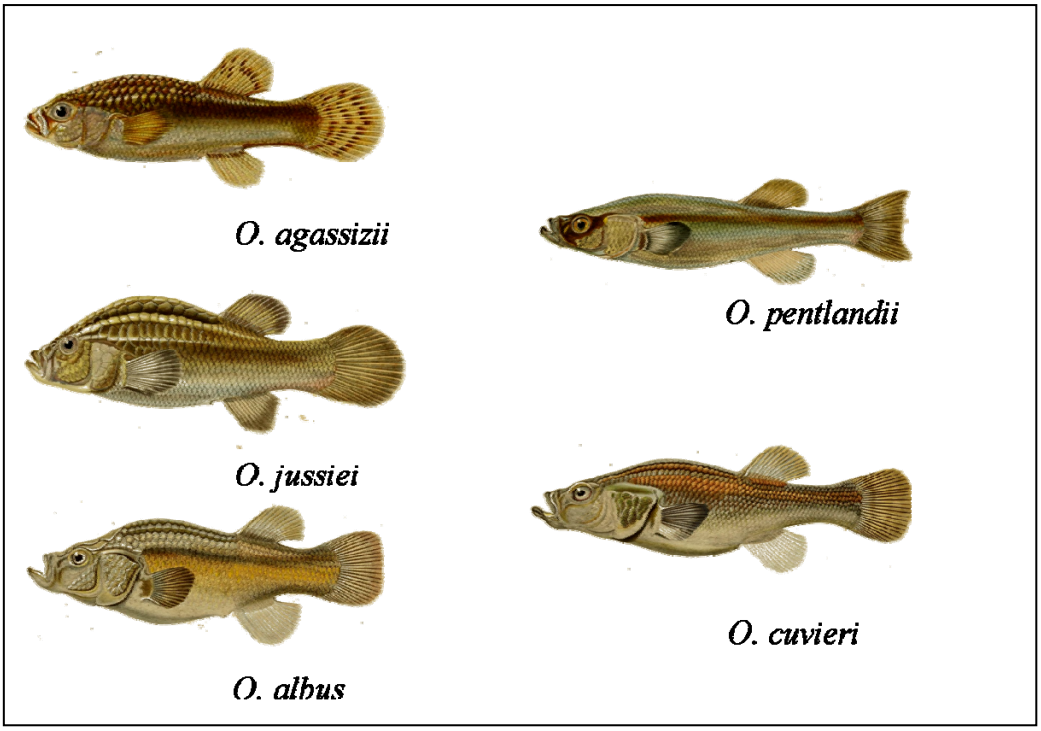


Figure 9. Some *Orestias* morpho-species from Lake Titicaca (modified from Valenciennes 1846).

Table 3. Spectrum of food categories (diet) and associated head morphology among four species co-occurring in Lake Titicaca (Maldonado *et al*. 2009).

Species	Zone	Diet	Associated head morphology
O. agassizii	mainly pelagic	Cladocera and algae	Short, blunt nose with a small mouth
O. albus	benthic	Preferentially amphipods but also fishes; the only piscivorous species after *O. cuvieri* (extinct)	Long snout with a large mouth
O. luteus	benthic	Preferentially mollusks	Larger posterior part of head and larger opercles
O. jussiei	littoral	Cladocera and algae, but able to forage on substratum	Short, blunt nose with a small mouth

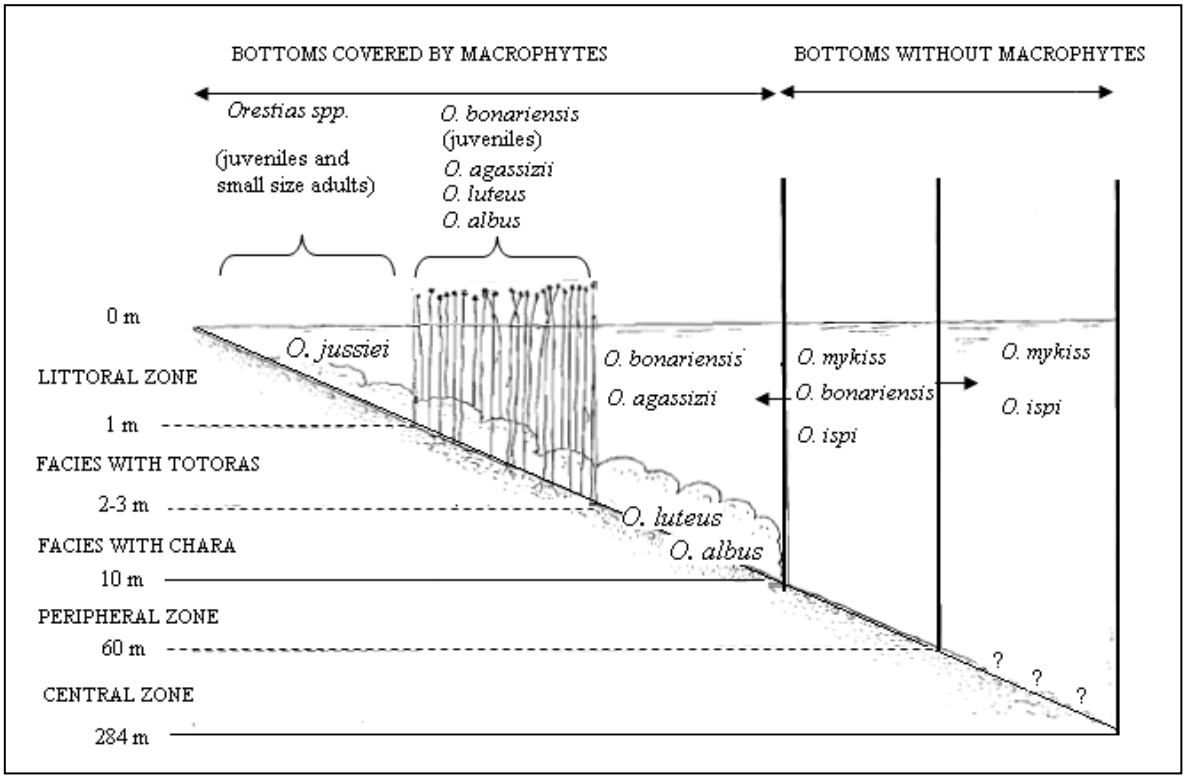


Figure 10. Habitats found in Lake Titicaca and some of the characteristic species, including the endemic *Orestias* species but also the silverside *Odontesthes bonariensis* and the rainbow trout *Oncorhynchus mykiss*, introduced in the lake between 1939 and 1956 (modified from Lauzanne 1992; Maldonado *et al*. 2009).

Orestias species are mostly small, rarely exceeding 5 cm in standard length, although they may reach almost 20 cm as in *O. pentlandii* (Lauzanne 1982; Parenti 1984b). They are characterized by sexual dimorphism in length and color. In a few cases, males may reach much larger body size than females, such as in *O. parinacotensis* (Parenti 1984b). A case of karyotypic dimorphism (i.e. heterochromy) was described in *O. laucaensis* (Parenti 1984b; Vila *et al.* 2010). Data regarding the basic aspects of breeding and spawning behavior of *Orestias* species —which are oviparous— are scarce. In Lake Titicaca, observations based on variation of the gonadosomatic index (GSI) have led to the conclusion that breeding takes place throughout the year without major variation among several species, including: *O. albus, O. forgeti, O. ispi, O. olivaceus, O. pentlandii, O. luteus* and *O. agassizii* (Lauzanne 1992).

Some species reach sexual maturity at small body size like for instance *O. minimus* (23.0 mm SL) and *O. minutus* (24.0 mm SL) (Parenti 1984b). Andean pupfishes produce demersal eggs variable in size (1.3 to 2.3 mm) that are deposited on submerged plants near the shore (Lauzanne 1992). The percentage of newborn males, starting from a sex ratio of 40-60%, decreases to reach very low values or even zero in large adults (Fig. 11). In addition, a potential case of 'protandrous hermaphroditism' (i.e. organisms that are born male and at some point in their lifespan turn into female) in *O. ispi* was suggested by Lauzanne (1992) as no females are found among the smallest individuals (length class: 40 to 45 mm) whereas the percentage of males falls very sharply to almost zero in the length class '70 to 75 mm' corresponding to the adult stage (Lauzanne 1992). In *O. tchernavini*, no males have ever been reported (Parenti 1984b; Lauzanne 1992), possibly representing a case of gynogenesis, a reproductive strategy common in other cyprinodontiform groups such as Poecilidae (Parenti 1984b).

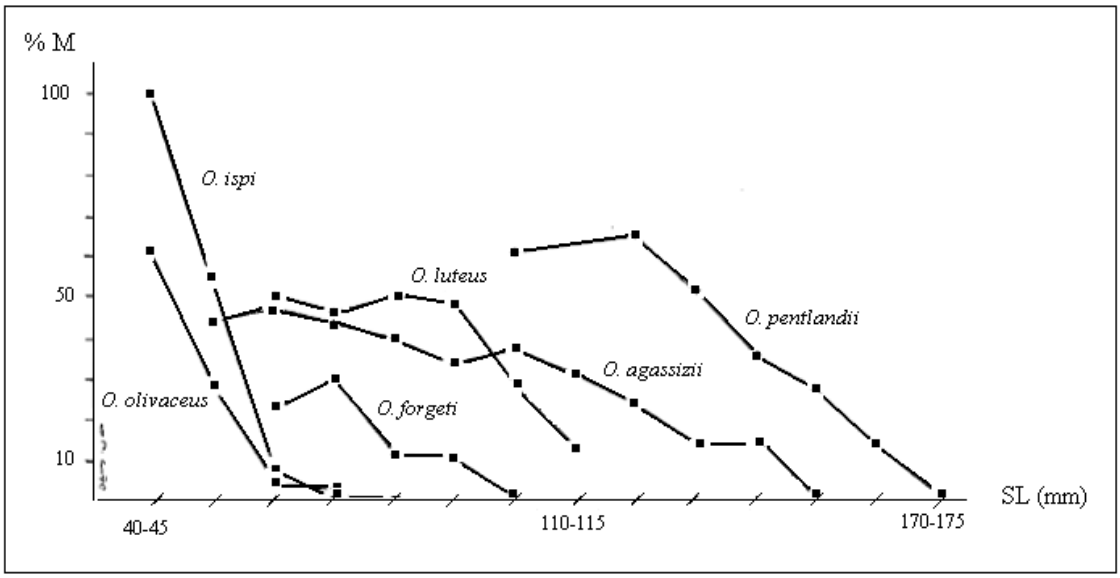


Figure 11. Changes in sex ratio (percentage of males) in relation to standard length (SL) in several species of *Orestias* (modified from Lauzanne 1992).

I.3.4 Conservation issues related to *Orestias*

In Lake Titicaca, where several species *Orestias* represent an important component of the fishery (Table 4) (Treviño *et al.* 1992; Capriles *et al.* 2008), continuous stock decrease (Vila *et al.* 2007b), rarefaction (e.g. *O. pentlandii*) and extinction (e.g. *O. cuvieri*) have been reported (Harrison & Stiassny 2004; CREO extinctions database http://creo.amnh.org/pdi.html; Van Damme *et al.* 2009) in association with deleterious factors such as overfishing (Orlove *et al.* 1992). Unfortunately, the impact of fishing pressure is difficult to assess accurately as fishing reports are scarce and most of them includes only some sectors of Lake Titicaca (e.g. Puno Bay) (Table 4). In addition, almost no data exists outside Lake Titicaca. Other factors includes: the introduction of invasive species (Loubens 1989; Lauzanne 1992; Orlove *et al.* 1992; Vila *et al.* 2007b), and environmental degradation through mining activities (Keller & Soto 1998; Vila *et al.* 2007a).

Table 4. Distribution of fisheries catches in Lake Titicaca (1980-1985) and Puno Bay (1989-1990) (Modified from Trevino *et al*. 1992).

Species	1980		1984-1985		1989-1990	
	(in tons)	%	(in tons)	%	(in tons)	%
O. agassizii	3417.94	54.01	1627.5	29	1.5	0.3
O. luteus	449.5	7.11	448.9	8	2.6	0.51
O. olivaceus	381.89	6.84	84.2	1.5	0.1	0.02
O. pentlandii	12.59	0.2	-	-	-	-
O. mulleri	14.28	0.22	28.1	0.5	0.3	0.06
O. cuvieri	21.28	0.34	-	-	-	-
O. ispi	17.05	0.27	-	-	0.1	-
Total genus *Orestias*	4313.53	68.19	2188.7	39	4.51	0.9
T. dispar	142.97	2.26	325.5	5.8	0.3	0.06
T. rivulatus	12.65	0.2	11.2	0.2	-	-
Total genus *Trichomycterus*	155.62	2.46	336.7	6	0.3	0.06
Total native species	4469.15	70.65	2525.4	45	9.62	1.91
Oncorhynchus mykiss	888.5	14.04	168.4	3	83.5	16.55
Odontesthes bonariensis	968.46	15.31	2918.2	52	411.3	81.54
Total introduced species	1856.96	29.35	3086.6	55	494.8	98.09
Total catches	6326.01	100	5612	100	504.42	100

Within the *agassizii* complex most of the species show a restricted distribution to very fragmented and fragile habitats that make them particularly vulnerable to human-induced perturbations and local extinctions (Parenti 1984b; Northcote 1992; Vila *et al.* 2007a; Vila *et al.* 2007b), such as small habitats in the southern region of the Altiplano of Bolivia and Chile (e.g. 'bofedales'). This region is subject to a negative water balance and water extraction for mining activities (Northcote 1992; Vila *et al.* 2007a; Vila *et al.* 2007b; Vila *et al.* 2013).

The synergism among these factors has led several *Orestias* species to become threatened and as consequence to be classified under several conservation categories (Table 5). However, for most of these species the conservation category has not been evaluated by the International Union for Conservation of Nature (IUCN) as many biological and ecological parameters remain unknown. This situation require urgent

conservation measures, especially since it may become exacerbated by the forthcoming effects of global climate change on the Altiplano (Kundzewicz *et al.* 2008).

In this regard, fish farming programs aiming at the rational production of fish and including growth rate and production control, has been developed in Lake Titicaca (mainly in the Peruvian part of the lake). In addition, since 1990 artificial spawning programs of native species, involving species such as *O. agassizii*, *O. luteus*, *O. ispi* and *O. albus* have been carried out because of the possibilities of these species becoming extinct and in order to restock the lake (Trevino *et al*. 1992). Unfortunately, no detailed information about the progress of these programs is available.

Table 5. Conservation status of *Orestias* species in the Libro Rojo de la fauna de vertebrados from Bolivia and Chile and in the IUCN. Categories: EX = extinct, CR = critically endangered, VU = vulnerable, DD = data deficient, LR/nt = lower risk/near threatened, NE = not evaluated.

Species	Libro Rojo	IUCN	Species	Libro Rojo	IUCN
O. cuvieri	EX	DD	*O. luteus*	VU	NE
O. pentlandii	CR	VU	*O. minimus*	VU	NE
O. albus	EN	NE	*O. minutus*	VU	NE
O. laucaensis	EN	LR/nt	*O. mooni*	VU	NE
O. parinacotensis	EN	DD	*O. mulleri*	VU	NE
O. chungarensis	EN	VU	*O. robustus*	VU	NE
O. agassizii	VU	NE	*O. taquiri*	VU	NE
O. crawfordi	VU	NE	*O. tchernavini*	VU	NE
O. forgeti	VU	NE	*O. tomcooni*	VU	NE
O. gilsoni	VU	NE	*O. tutuni*	VU	NE
O. gracilis	VU	NE	*O. uruni*	VU	NE
O. imarpe	VU	NE	*O. silustani*	VU	VU
O. incae	VU	NE	*O. ctenolepis*	VU	VU

I.4 The study system – part II: the Altiplano and Lake Titicaca

The freshwater systems on the Altiplano are endorrehic and are subjected to extreme conditions typical of high mountainous climate such as total irradiance, low temperatures and dry conditions. The high evaporation and geology produce lake waters with a wide salinity range showing an average of 1 g 1^{-1} and highly variable pH conductivity values ranging between 8.0 to 90 µmhos cm^{-1}. The mean annual temperature in this region is low and range between 5ºC in July and 10ºC in December (Vila *et al*. 2007). As a consequence the taxonomic diversity level of the regional fauna is relatively low compared to neighboring hydrographic systems such as in the Amazonia (Hoorn *et al.* 2010). For instance, only one species of *Coelenterata* belonging to the genus *Hydra* occurs in Lake Titicaca. Similarly, triclads (turbellarian flatworms) are represented by a single, polytypic species (*Euplanaria dorotocephala*), showing a great variability in color and shape throughout its widespread distribution (Dejoux 1992). Nevertheless, and despite such harsh climatic conditions, several cases of putative species flocks have been described within Lake Titicaca, notably in the amphipod *Hyallela* (Gonzalez & Watling 2003; Väinölä *et al.* 2008) and the microgastropod *Heleobia* (Hershler & Thompson 1992) (see Table 1). Unfortunately, as with *Orestias*, the phylogenetic and biogeographical relationships of such groups are still poorly understood.

I.4.1 Formation of the Altiplano and succession of paleolakes at the origin of Lake Titicaca

The Altiplano is an inter-montane endorheic plateau lying between the western and eastern slopes of the southern region of the central Andes from Peru, to Bolivia, Chile and Argentina. It spans 200,000 km^2 from 14 °S to 23 °S (Fig. 12) with an altitudinal

range of 3600-4500 m, and represents the southern part of the inter-Andean basin (Lavenu 1992; Wirrmann & Mourguiart 1995; Vila *et al.* 2007b). At present, this region is under the influence of an arid to semi-arid climate with a north-south metereological gradient and a southward water flow (Wirrmann & Mourguiart 1995). The hydrological balance depends on the rate of precipitation and evaporation, the latter being approximately 90% (Carmouze *et al.* 1981), with average annual rainfalls dropping from 900 mm in the north to less than 100 mm in the south. The north-central region of the Altiplano is mostly characterized by the basin including Lake Titicaca. The latter is composed of Lake Titicaca, Rio Desaguadero, Lake Poopo and Salar de Coipasa, and constitute the 'TDPS' system, with a total coverage of approximately 140,000 km^2 (Lavenu 1992).

Lake Titicaca is the largest lake of South America (8560 km^2) and the highest navigable lake in the world (3808 m). Its southern outlet Rio Desaguadero (320 km long; 3804 m) drains southward into Lake Popoo (2530 km^2; 3685 m) (Fig. 12). The latter is characterized by dryer climatic conditions (mean annual rainfall: 390 mm), but in especially rainy periods may overflow into Salar de Coipasa (2500 km^2; 3657 m). During rainy years, this salt pan can be filled with shallow waters (<4 m) originating from Rio Lauca (situated in the western border of Bolivia with North Chile, near the Salar de Coipasa) and Rio Grande de Lipez (situated in South Bolivia). The southern region of the Altiplano (18-21° S) is represented by a volcanic, arid and fragmented landscape, characterized by a large number of bofedales (e.g. Sur Lipez region), lakes and salt pans. Salar de Uyuni, south of Salar de Coipasa, represents the largest salt pan in the world (12,000 km^2; 3653 m) and is mainly fed by Rio Grande de Lipez. Both salt pans sporadically communicate when precipitation levels are high (Roche *et al.* 1992; Sylvestre *et al.* 1999; Argollo & Mourguiart 2000; Fornari *et al.* 2001). Other salt pans

of smaller extension are also present in the Chilean Altiplano, including Salar de Ascotán (243 km^2; 3723 m), Salar de Carcote (108 km^2; 3700 m), and Salar de Huasco (51 km^2; 4550 m). In this zone, contemporary connections among freshwater systems may exist, and are almost entirely subterranean (Vila 1975; Niemeyer & Cereceda 1984).

The inter-Andean basin as it stands has been dated to the end of the Miocene. Decoud *et al*. (2013) proposed that the main period of uplift of the Central Andes started approximately 25 Myr to stop around 10 Mya (see also Ghosh *et al.* 2006). Since the early Quaternary (from 1.6 Mya to 14 kya), the Altiplano has been occupied by large paleolakes. The study of deposits and shorelines allowed the identification of several episodes of paleolakes and glacial stages (Lavenu *et al.* 1984; Lavenu 1992; Sylvestre *et al.* 1999; Fornari *et al.* 2001; Placzek *et al.* 2006). These paleolakes were subject to alternated episodes of expansion and desiccation due to cyclic climatic changes (Wirrmann & Mourguiart 1995). The temporal succession of the paleolakes was marked by a decrease in water levels, to reach the current state of Lake Titicaca (Fig. 13 and 14). These paleolakes are:

i) Paleolake Mataro, the largest recorded paleolake expansion, overlapping much of the Altiplano. This paleolake culminated at 3950 m (ca. 140 m higher than present Lake Titicaca) and reached its maximum extension after the Calvario formation (1.5-1.6 Mya).

ii) Paleolake Cabana, during the Kaluyo/Sorata interglacial episode (1.1 – 1.0 Mya), reaching a maximum elevation of 3900 m.

iii) Paleolake Ballivian, during the Sorata/Choqueyapu I interglacial (0.6-0.5 Mya), reaching a level of 3860 m. During the Ballivian lacustrine event, the main TDPS lakes were larger and connected to each other by straits.

iv) Paleolake Minchin, corresponding to the Choqueyapu I/II (t2), with an altitudinal level of 3825 m. This paleolake, divided into two water bodies by the Ulloma-Callapa sill (located in the Rio Desaguadero basin, see Fig. 13), was very reduced in extent north of Lake Titicaca, but had a larger area in southern Altiplano, covering the Poopo and Coipasa-Uyuni basins.

v) Paleolake Tauca, during the postglacial Choqueyapu II phase (18-14.5 kya), with a level of 3815 m. This lake was also divided into two water bodies by the Ulloma-Callapa sill (see Fig.13), but had a more reduced area than paleolake Minchin. Lake Tauca reached 5 m above the present water level of Lake Titicaca.

During the Late Pleistocene and Holocene, a phase of paleolake succession from Ballivian to Tauca episodically integrated into a single, large basin, the drainage basins of the southern Altiplano, including Poopo, Coipasa and Uyuni (Blodgett *et al.* 1997), thus suggesting several episodes of geologically 'recent' inter-connections among the current basins of the Altiplano. Lake-level decrease due to evaporation and lowering of mean annual rainfall (sometimes resulting in a rapid 100 m drop in water level) persisted during the Holocene, with the consequence of drastically increasing salinity levels in Lake Titicaca, Lake Poopo and other lakes in the region of Sur Lipez (Wirrmann & Mourguiart 1995; Cross *et al.* 2001; Placzek *et al.* 2006).

It is generally assumed that such paleohydrological events have shaped the evolutionary history of the Altiplano aquatic biota, with repeated cycles of lake level variation promoting dispersal and vicariance, and desiccation and associated fluctuations in salinity resulting in local extinctions and habitat fragmentation (Kroll *et al.* 2012; Vila *et al.* 2013).

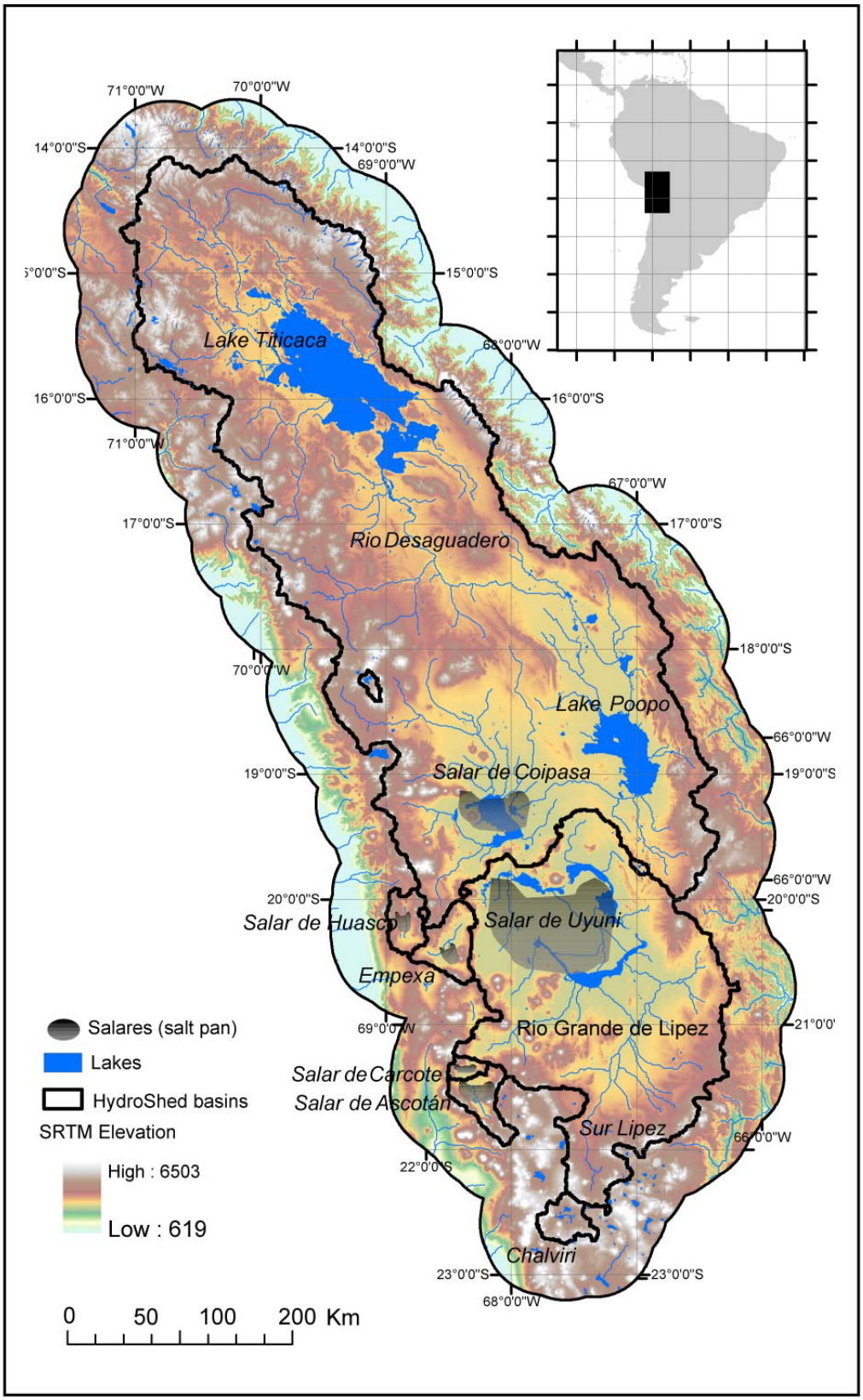


Figure 12. The Andean Altiplano, delimited by basins. The 'TDPS' hydrographic system includes Lake **T**iticaca, Rio **D**esaguadero, Lake **P**oopo and Salar de **C**oipasa.

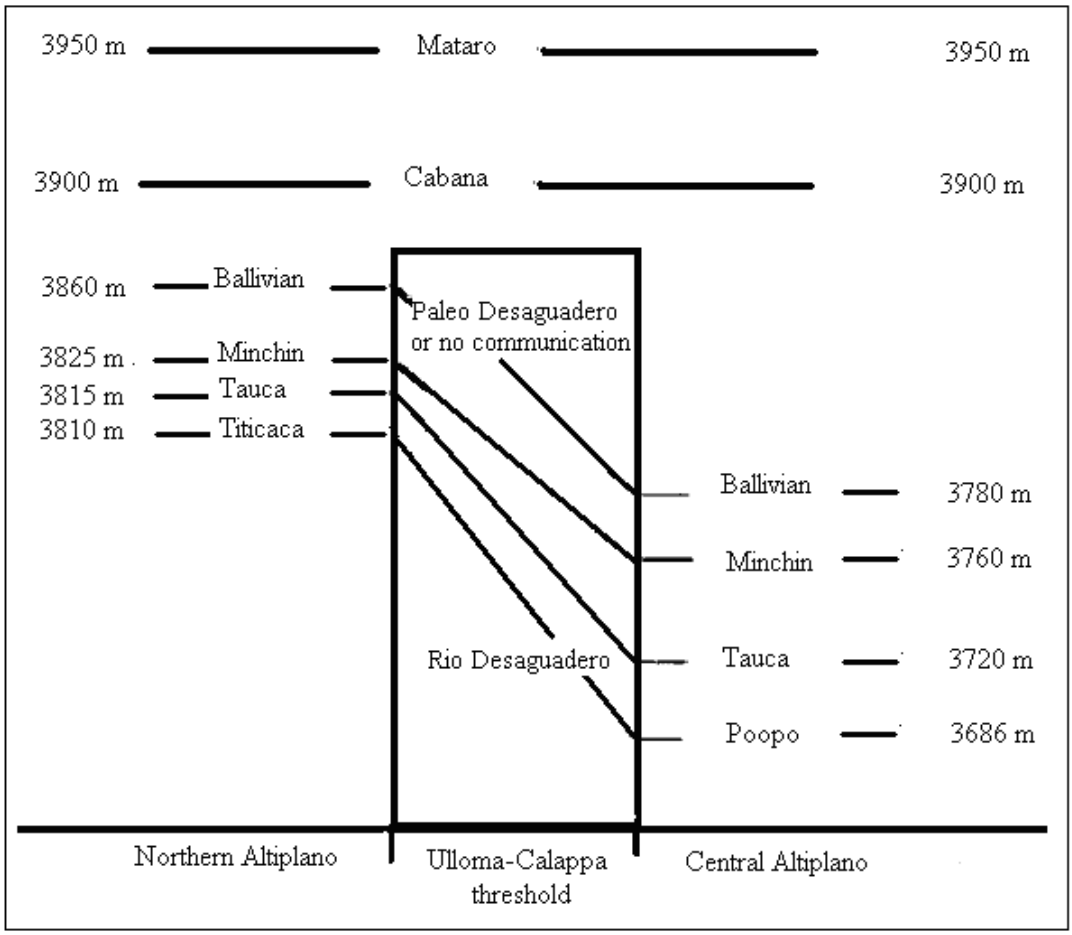


Figure 13. Succession of the Altiplano paleolakes (from top to bottom) until the current Lake Titicaca (see §I.4.1. for geochronological scale). Escara and Poopo represent the southern basins of Paleolake Minchin and Lake Titicaca, respectively (modified from Lavenu 1992).

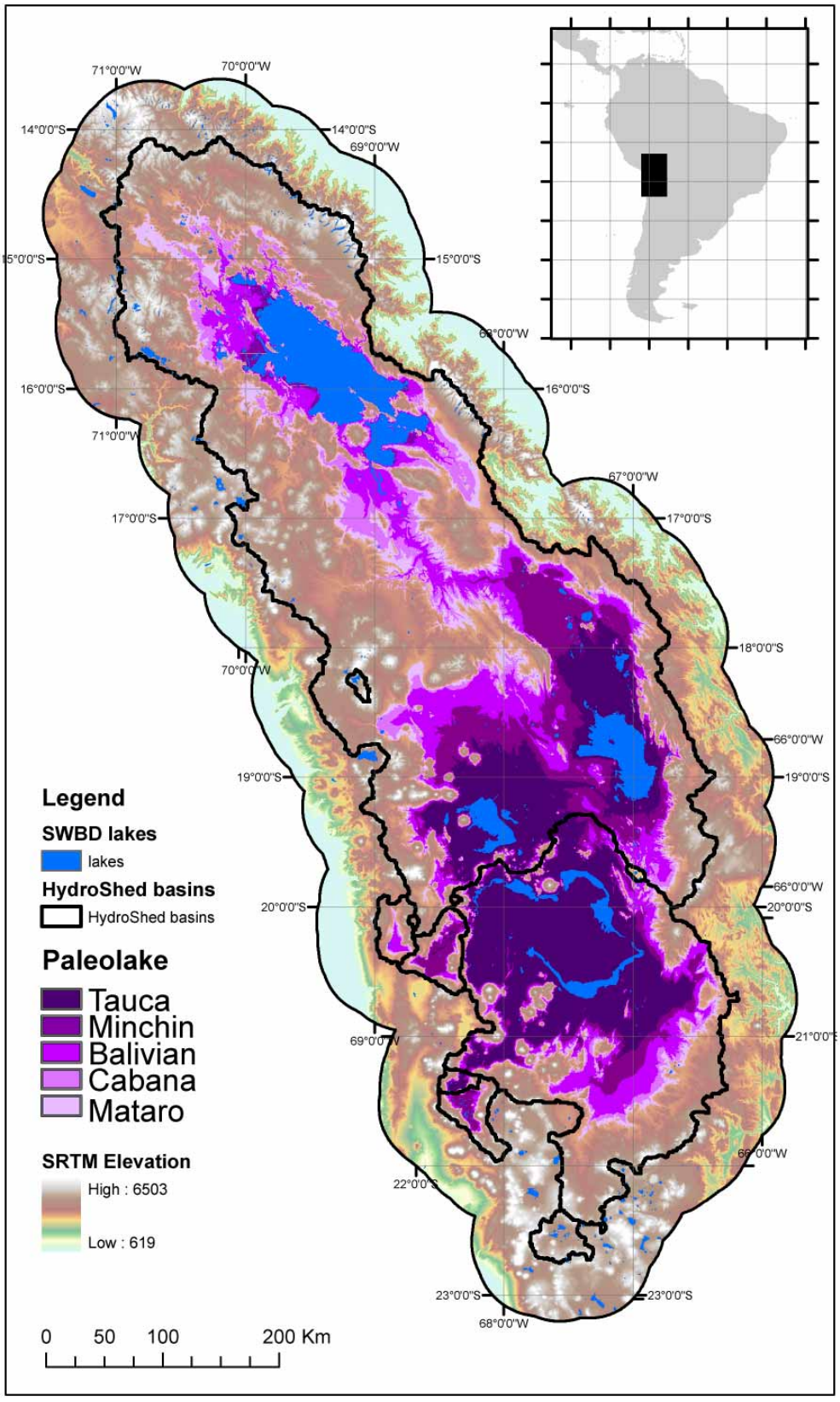


Figure 14. Maximal lake level extension of the Altiplano paleolakes.

1.5 Context and objectives.

1.5.1 Why study the *agassizii* complex when questioning radiation?

With regard to other freshwater teleosteans of the region, the genus *Orestias* is the sole to be endemic to the inter-Andean basin and is by far the most speciose. Since the description of Tchernavin (1944), the genus has been referred to a species flock (Villwock 1962; Kosswig & Villwock 1964) because of its great ecological and morphological diversity. Knowledge on the factors and mechanisms underlying the morphological diversification and speciation of *Orestias* are very fragmentary. Some authors have suggested that the genus may have gone through an important diversification phase during the Pleistocene (Parenti 1981; Parenti 1984b; Lüssen *et al.* 2003), likely processing as an event of adaptive radiation within Lake Titicaca among different complexes (Villwock 1986; Maldonado *et al.* 2009). To date no exhaustive molecular-based investigations have been conducted —apart from a taxonomically partial analysis made by Lüssen *et al*. (2003) that joined similar conclusions.

Association between a particular morphology (e.g. head, jaw and gillraker shape in teleosteans) and a specific niche (e.g. planktivores *vs*. piscivores) is recognized as an important signature of adaptive radiation (Schluter 2000; Danley & Kocher 2001; Hulsey *et al.* 2010; Von Rintelen *et al.* 2010). In this sense, Maldonado *et al*. (2009) provided evidence for the association between foraging and feeding strategies and particular morphological adaptations in several *Orestias* species inhabiting in Lake Titicaca, that may support adaptive radiation within the genus.

In line with the definition of a species flock (see §I.1.), I opt here for considering the *agassizii* complex as a species flock on itself (contra Parenti 1984a), because (i) it is monophyletic, as redefined by Lüssen *et al*. (2003) and in this work (Article 1), (ii) it is endemic to a geomorphologically delimited area (inter-Andean basin), (iii) it has a great

number of morpho-species (more than the other complexes cumulated), and (iv) it shows a fair level of eco-morphological diversification among the described species. Since the *agassizii* complex is not endemic to a given lake, my concept of species flock applied to this particular taxon rejoins the 'riverine species flock' model of Sullivan *et al*. (2002). Although here, it would be more appropriate to talk about a 'lacustrine-riverine species flock'.

The particularity of the *agassizii* complex is that its radiation-like diversification may have occurred in two different evolutionary contexts, including (i) with interspecific competition and no physical barrier against gene flow in Lake Titicaca, and (ii) through dispersion and without inter-specific competition across the inter-Andean basin (i.e. outside Lake Titicaca). This suggests that adaptive radiation is not the only potential evolutionary framework explaining the diversification of *Orestias*. More probably, a combination of processes including ecological speciation within Lake Titicaca and —for the *agassizii* complex only— fixation of mutations in separate populations adapting to similar selection pressures outside the lake ("mutation-order speciation" Schluter 2009) both contributed to the current species flock-like pattern of the complex.

I.5.2 Objectives

The general aim of my work was to elucidate the mechanisms and factors underlying the diversification and speciation in the genus *Orestias*, especially within the most speciose and widespread *agassizii* complex, excluding the '*luteus*' group after Lüssen *et al*. (2003). As a corollary to this aim, I am expecting to reassess taxonomic delineations among complexes and species of *Orestias* using molecular phylogenetic approaches, in order to accurately delineate the phylogenetic position of the *agassizii* complex. In a

group with such ‘fuzzy’ morpho and/ or osteological delimitations, phylogenies generated from characters independent of morphology are of great interest. In addition, I will try to assess the effect of hybridization on the morphotypic variation within the *agassizii* complex (*O. luteus x O. agassizii* hybrids). I will discuss the possible impact of hybridization on long-standing taxonomic difficulties and will question the role of hybrids in promoting the morphological diversification of the complex. Finally, I will conduct population genetic study on regional scale focused on the *agassizii* complex to question the potential role of Pleistocene paleolake fluctuations on the diversification processes shaping the radiation of the complex across the Altiplano, notably outside Lake Titicaca.

The synergy between these objectives is consolidated through a multi-scale approach using nucleotide sequences (mitochondrial and nuclear DNA), microsatellites, morphology (morphometrics and meristic characters), ‘environmental’ data (altitude, hydrographic distances) and methods of analysis covering the fields of phylogeny, phylogeography and population genetics.

I.5.2.1 Specific objectives:

- Objective 1. Given that a limited number of molecular markers (mitochondrial DNA) have been reported in the literature for *Orestias*, the first objective of my work focused on the isolation and characterization of independent, polymorphic molecular markers that would be useful in our inferences at different analytical scales such as taxonomic delimitation, assessment of hybridization, phylogeny, phylogeography and population genetics (Article 2; Table 4).

- Objective 2. My second objective was to clarify the taxonomic delimitations among *Orestias* complexes and morpho-species using an ancient DNA phylogenetic approach

including type specimens, so I could define on the monophyly criterion the evolutionary lineage corresponding to the *agassizii* complex (Article 1).

- Objective 3. My third objective was to assess the incidence of natural hybridization between two of the most representative species within the genus, *O. agassizii* and *O. luteus*, in order to provide an integrative diagnostic tool for hybrids and discuss the potential impact of hybridization on the evolution and conservation of *Orestias* in general, and *O. agassizii* in particular (Article 3).

- Objective 4. My fourth objective was to assess the genetic structure of the *agassizii* complex at a population genetic level in order to question the role of Pleistocene paleolake fluctuations on the diversification processes shaping the radiation of the complex across the Altiplano, notably outside Lake Titicaca, and eventually to propose a scenario of diversification for the complex (Article 4).

I.6 Material and methods - Synopsis

My PhD thesis benefitted from two international collaborations that enabled the team "Biodiversité et Macroécologie" (led by Bernard Hugueny; IRD) and various collaborators to support the sampling effort in Bolivia and Chile (Fig. 15 and Table 7), including i) ANR FISHLOSS 2008-2013 (PI: Bernard Hugueny, IRD; chief collaborator: Carla Ibañez, Universidad Mayor de San Andres La Paz, Bolivia), and ii) ECOS-sud 2011-2013, with Universidad de Chile (chief collaborator: Marco Mendez). Altogether, approximately 2650 specimens were collected and tissue-sampled (Peruvian-Bolivian samples n=2470 - 46 localities; Chilean samples n=177 - 6 localities). Bolivian-Peruvian specimens were photographed and measured for 17 standard morphometric variables as described in Article 3. In addition, sex, gonadic state and weight were recorded. Whenever possible, stomach content was preserved for

future studies on trophic niches. To achieve the objectives of my work, a subsample of 1325 specimens representing the four traditional complexes and the *luteus* group were selected for genetic and morphometric analyses. Given the difficulties in identifying morpho-species within *Orestias*, we proceeded to a phylogenetic classification of the nucleotide sequences obtained from the specimens (see Article 1). At the complex (clade) level, taxonomic assignments and sample numbers were as follows: (i) *agassizii* complex - from Bolivia *N*=917, from Chile *N*=177; (ii) *luteus* group *N*=154; (iii) *cuvieri* complex *N*=26; (iv) *gilsoni* complex *N*=51. The *mulleri* complex was polyphyletic. In addition, 16 type specimens from the Muséum National d'Histoire Naturelle of Paris were sampled for genetic analysis, including representatives of all the *Orestias* complexes and the extinct *O. cuvieri* (Article 1).

A series of molecular markers potentially useful to unravel the genetic relationships within *Orestias* were identified from the literature and tested on our samples (Article 2; Table 6). We were especially looking for markers that could be polymorphic at the intra-complex level (*agassizii*) but that could also be useful to reconstruct the phylogenetic relationships among *Orestias* complexes. PCR amplification and levels of polymorphism were assessed in a total of eighty specimens representing the four *Orestias* complexes as defined by Parenti (1984b). I also developed and tested 10 microsatellite loci usable for two complexes (*agassizii* and *luteus*) and that could also be used to genotype the other complexes (Article 2). In total, and after the final selection of the most useful loci, I generated 2197 mitochondrial sequences (cytochrome *b* and control region), 633 nuclear sequences (rhodopsin, transmembrane domains I – VI) and 1113 genotypes that were used across the different analytical scales of my work.

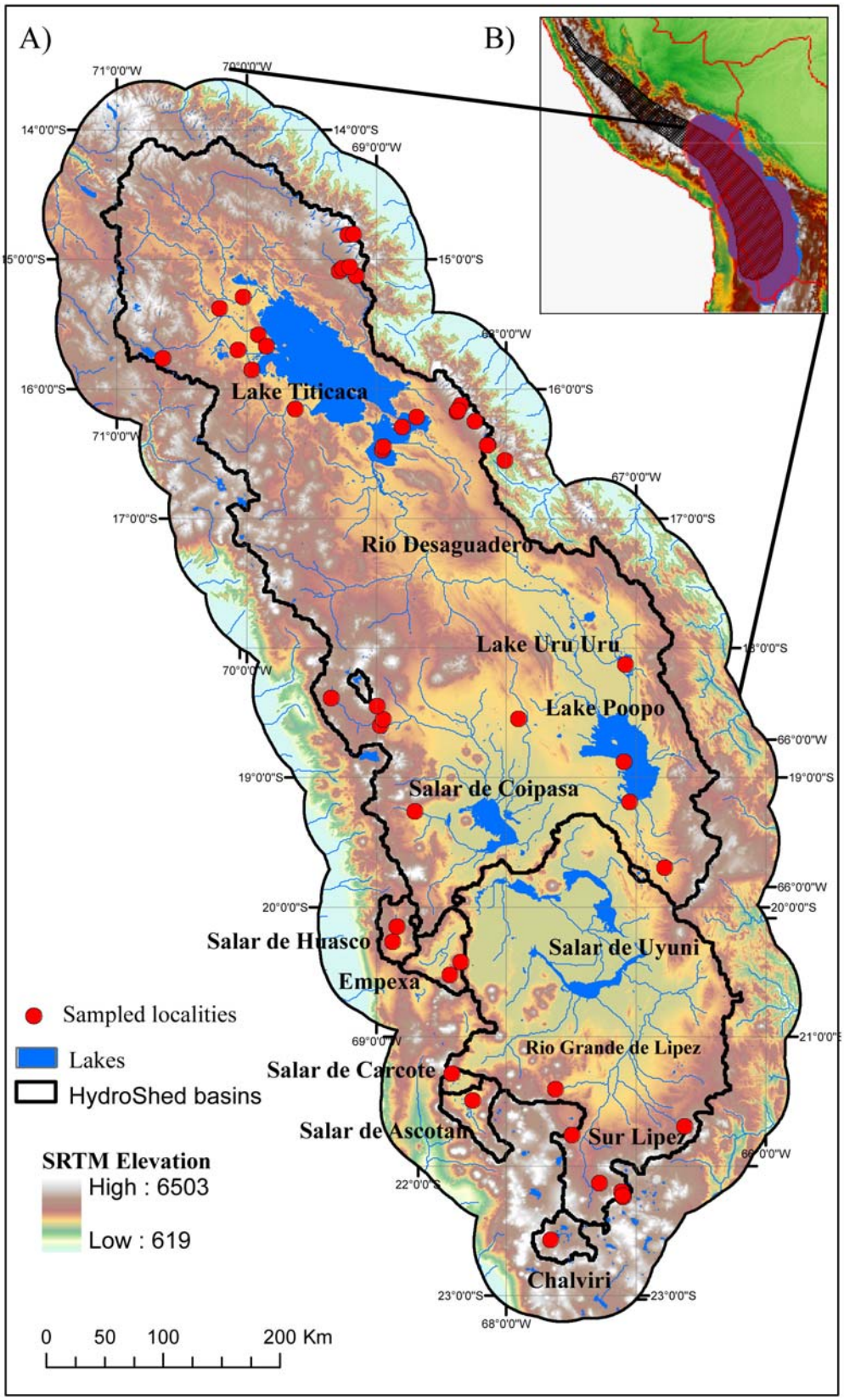


Figure 15. Sampling coverage across the Altiplano for *Orestias* species and populations considered in the study, with (A) the sampled sites (red circles), and (B) the distribution range of the genus (black area) and the sampled area (purple).

Table 6. Molecular markers tested for the phylogeny, phylogeography and population genetics of *Orestias* taxa. Markers retained for my PhD project are indicated with an asterisk. NA: Not assessed.

Molecular marker	Amplified fragment length (bp)	Informative characters / Issues	Analytic level	Ref
Mitochondrial DNA:				
Control region *	398	Yes	Phylogeny	(Falk *et al.* 2003)
Cytochrome *b* *	1102	Yes	Phylogeography Hybrid detection	Machordom & Doadrio (2001)
COI	670	Low level of polymorphism within complexes	NA	(Ward *et al.* 2005)
Nuclear DNA:				
Ribosomal protein gene, S71	ca. 900	Yes, multiple copies	Phylogeny Phylogeography	(Chow & Hazama 1998)
Ribosomal protein gene, S72	ca. 250	Yes, multiple copies	Hybrid detection	Id.
Creatine Kinase, intron 7	165	No	NA	(Quattro & Jones 1999)
Lactate Dehydrogenase, intron 6	171	Two polymorphic sites among complexes	NA	Id.
Myosine, intron 6	70	No	NA	Id.
[EPICS] 5537BE1	ca. 900	Yes (9 synapomorphic sites discriminating between *agassizii* and the remaining complexes) Not retained because redundant with the signal of Rhodopsin	NA	(Li *et al.* 2007)
[EPICS] 04174E20	ca. 1000	Low polymorphism and PCR amplification problematic	NA	Id.
Recombination Activation Gene	275	No, low level of polymorphism and multiple copies	NA	(Carson & Dowling 2006)
Rhodopsin, transmembrane domains I - VI *	690	Yes (7 polymorphic sites discriminating between *agassizii* and the remaining complexes) Bi-allelic	Phylogeny Hybrid detection	(Chen *et al.* 2003) Articles 1 and 3
Microsatellites: 10 loci * (isolated from *O. agassizii*)*	140-400	Yes	Population genetics Hybrid detection	Article 2

Table 7. Sampled localities that were considered in the genetic analyses, with their general characteristics.

Country	Locality	Habitat	Altitude (m)	Latitude (S)	Longitude (W)
Bolivia	Titicaca, Huatajata	lake	3788	-16.2137	-68.6901
Peru	Titicaca, Puno	lake	3791	-15.8489	-69.9641
Peru	Titicaca, Coata	lake	3797	-15.5784	-69.9098
Peru	Titicaca, Capachica	lake	3797	-15.6654	-69.8506
Bolivia	Titicaca, San Jose	lake	3789	-16.4675	-68.9573
Bolivia	Titicaca, Santa Rosa	lake	3795	-16.4447	-68.9493
Bolivia	Titicaca, Isla Taquiri	lake	3807	-16.2899	-68.8019
Peru	Juliaca	lake	3830	-15.3744	-70.2108
Peru	Ramis	river	3812	-15.2885	-70.0268
Peru	Zapatilla	river	3843	-16.1523	-69.6279
Peru	Ilpa	river	3819	-15.6953	-70.0684
Peru	Saracocha	lake	4135	-15.7623	-70.6465
Bolivia	Catantica	bofedal	4657	-14.8072	-69.2233
Bolivia	Catantica Alto	bofedal	4823	-14.8029	-69.1799
Bolivia	Chojña Khota	lake	4431	-15.1246	-69.1575
Bolivia	Suches	river	4328	-15.0864	-69.2864
Bolivia	Ulla Ulla	bofedal	4335	-15.0633	-69.2610
Bolivia	Cañuhma Janka Khota	lake	4450	-15.0548	-69.2070
Bolivia	Hichu Khota	lake	4335	-16.1742	-68.3800
Bolivia	Khotia	lake	4453	-16.1231	-68.3514
Bolivia	Khara Khota	lake	4335	-16.1624	-68.3725
Bolivia	Tuni	lake	4436	-16.2465	-68.2419
Bolivia	Siete Lagunas (2)	lake	4273	-16.4281	-68.1433
Bolivia	Siete Lagunas (4)	lake	4241	-16.4319	-68.1472
Bolivia	Siete Lagunas (5)	lake	4236	-16.4322	-68.1475
Bolivia	Siete Lagunas (6)	lake	4219	-16.4329	-68.1487
Bolivia	Huni	lake (translocation)	3885	-16.5481	-68.0125
Bolivia	UruUru	lake	3682	-18.1233	-67.0841
Bolivia	Poopo	lake	3673	-18.8760	-67.0926

Table 7. (Continued).

Bolivia	Pampa Aullgas	river	3665	-19.1844	-67.0479
Bolivia	Mulato	river	3812	-19.6906	-66.7794
Bolivia	P. Opoqueri	lake (artificial)	3767	-18.5435	-67.9054
Bolivia	Lauca	river	3868	-18.5937	-68.9707
Bolivia	Macaya	lake	3855	-18.5454	-68.9504
Bolivia	Mogachi	river	4628	-18.4426	-68.9957
Bolivia	Empexa	river	3781	-20.5169	-68.4341
Bolivia	Barras	river	3971	-20.4202	-68.3500
Bolivia	Villa Mar	river	4029	-21.7530	-67.4929
Bolivia	San Pablo de Lípez	river	4227	-21.6881	-66.6261
Bolivia	Sol de Mañana	river	4187	-22.1278	-67.2823
Bolivia	Villa Alota	bofedal	3820	-21.4003	-67.6224
Bolivia	Chipapa	lake	4461	-22.1934	-67.1112
Bolivia	Celeste	bofedal	4481	-22.2327	-67.1022
Bolivia	Celeste	bofedal	4461	-22.2224	-67.1053
Bolivia	Chalviri	bofedal	4409	-22.5653	-67.6568
Peru	Lagunillas	lake	4100	-15.4105	-70.4681
Chile	Ascotán	salt pan	3722	-21.4892	-68.2615
Chile	Carcote	salt pan	4613	-21.2795	-68.4206
Chile	Collacagua	river	3835	-20.1466	-68.8436
Chile	Huasco	salt pan	3789	-20.2628	-68.8754
Chile	Isluga	river	3767	-19.2564	-68.7063
Chile	Lauca	river	4272	-18.3807	-69.3489

Chapter II

- **Articles**

II.1 List of articles

My PhD work resulted so far in two accepted publications, one submitted manuscript and one manuscript in preparation (Table 6):

- Article 1: Yareli Esquer Garrigos, Bernard Hugueny, Kellie Koerner, Carla Ibañez, Celine Bonillo, Patrice Pruvost, Romain Causse, Corinne Cruad & Philippe Gaubert (2013) Non-invasive ancient DNA protocol for fluid-preserved specimens and phylogenetic systematics of the genus *Orestias* (Teleostei: Cyprinodontidae). *Zootaxa*, 3640, 379-394.

- Article 2: Yareli Esquer Garrigos, Josie Lambourdière, Carla Ibañez & Philippe Gaubert. (2011) Characterization of ten polymorphic microsatellite loci in the Andean pupfish *Orestias agassizii*, with cross-amplification in the sympatric *O. luteus*. Conservation Genetics Resources, 3, 117-119.

- Article 3: Yareli Esquer Garrigos, Bernard Hugueny, Carla Ibañez, Kellie Koerner, Claudia Zepita, Josie Lambourdiere, Arnaud Couloux & Philippe Gaubert. Hybridization within an adaptive radiation framework: evolutionary and conservation implications for two species of Andean pupfishes (*Orestias* spp., Teleostei, Cyprinodontidae). Molecular Ecology (submitted).

- Article 4: Yareli Esquer Garrigos, Bernard Hugueny, Carla Ibañez, Marco Méndez, Pamela Morales and Philippe Gaubert. Phylogeographic analysis of the *agassizii* complex (genus *Orestias*: Cyprinodontidae): a rapid diversification following dispersal across the Andean Altiplano? (in preparation).

Table 1. Brief description of the specific objectives for each article, specifying data sets and analytical scales.

Objectives	Data Sets	Analytic Scales	Publications (and status)
To assess the validity of the *Orestias* complexes proposed by Parenti (1984), to infer their phylogenetic relationships, and to test the validity of the diagnostic morphological characters used to describe the complexes	-Representatives of all *Orestias* complexes, including freshly collected and type specimens. -Sample size: *N*=44 specimens representing 11 species from freshly collected samples and 11 species as type specimens. -Type of data: mtDNA sequences (control region) and nDNA sequences (rhodopsin), morphometrics and meristic characters. -Specific technique: Ancient DNA.	Phylogeny	Article 1 (published in Zootaxa)
To characterize microsatellite markers for population genetic analyses on two *Orestias* complexes supposed to hybridize	-Representatives of both *O. agassizii* and *O. luteus*. -Sample size: *N*=46 . -Type of data: Microsatellites	Population genetics	Article 2 (published in Conservation Genetics Resources)
To assess the incidence of hybridization between *O. agassizii* and *O. luteus* and to characterize hybrids using molecular and morphological approaches, and to discuss the impact of hybridization on the diversification and conservation of *Orestias* species	-Specimens of *O. agassizii*, *O. luteus* and putative hybrids. -Sample size: *N*=175. 108 specimens of *O. agassizii*, 50 specimens of *O. luteus* and 17 putative hybrids. -Type of data: mtDNA sequences (control region and cytochrome *b*), nDNA sequences (rhodopsin), 10 microsatellite loci, morphometrics and meristic characters.	Phylogeny, Population genetics	Article 3 (submitted to Molecular Ecology)
To assess the genetic structure within the *agassizii* complex, and to test the impact of Pleistocene hydrographic dynamics on the diversification of the complex	-Specimens of the *agassizii* complex. Sample size: *N*=996 specimens collected in the Altiplano, from Bolivia, Peru and Chile. -Type of data: mtDNA sequences (control region and cytochrome *b*), nDNA sequences (rhodopsin), 9 microsatellite loci.	Population genetics Phylogeography	Article 4 (in preparation)

II.2 Article 1

Yareli Esquer Garrigos, Bernard Hugueny, Kellie Koerner, Carla Ibañez, Celine Bonillo, Patrice Pruvost, Romain Causse, Corinne Cruad & Philippe Gaubert (2013) **Non-invasive ancient DNA protocol for fluid-preserved specimens and phylogenetic systematics of the genus *Orestias* (Teleostei: Cyprinodontidae)**. Zootaxa, 3640, 379-394.

Zootaxa 3640 (3): 373–394
www.mapress.com/zootaxa/

Article

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

http://dx.doi.org/10.11646/zootaxa.3640.3.3
http://zoobank.org/urn:lsid:zoobank.org:pub:5D86BB42-FB1A-4873-9349-B5B225060F65

Non-invasive ancient DNA protocol for fluid-preserved specimens and phylogenetic systematics of the genus *Orestias* (Teleostei: Cyprinodontidae)

YARELI ESQUER GARRIGOS[1,6], BERNARD HUGUENY[1], KELLIE KOERNER[1,2], CARLA IBAÑEZ[1,3], CELINE BONILLO[4], PATRICE PRUVOST[1], ROMAIN CAUSSE[1], CORINNE CRUAUD[5] & PHILIPPE GAUBERT[1]

[1]*UMR BOREA, Département Milieux et Peuplements Aquatiques MNHN - CNRS 7208 - IRD 207 - UPMC, Muséum National d'Histoire Naturelle, 43 rue Cuvier, 75231 Paris, France*
[2]*Current address: Charter School of the Dunes, 860 N Lake Street, Gary, IN 46403, USA*
[3]*Instituto de Ecología – Unidad de Limnología, Universidad Mayor de San Andrés, Campus Universitario Cota cota C.27 s/n, La Paz, Bolivia*
[4]*UMS Outils et méthodes de la systématique intégrative CNRS 2700, Département Systématique & Evolution, Muséum national d'Histoire naturelle, 57 rue Cuvier, 75005 Paris, France*
[5]*Genoscope. Centre National de Sequençage, 2 rue Gaston Crémieux, CP5706, 91057 Evry Cedex, France*
[6]*Corresponding author. E-mail: esquer@mnhn.fr*

Abstract

Specimens stored in museum collections represent a crucial source of morphological and genetic information, notably for taxonomically problematic groups and extinct taxa. Although fluid-preserved specimens of groups such as teleosts may constitute an almost infinite source of DNA, few ancient DNA protocols have been applied to such material. In this study, we describe a non-invasive Guanidine-based (GuSCN) ancient DNA extraction protocol adapted to fluid-preserved specimens that we use to re-assess the systematics of the genus *Orestias* (Cyprinodontidae: Teleostei). The latter regroups pupfishes endemic to the inter-Andean basin that have been considered as a 'species flock', and for which the morphology-based taxonomic delimitations have been hotly debated. We extracted DNA from the type specimens of *Orestias* kept at the Muséum National d'Histoire Naturelle of Paris, France, including the extinct species *O. cuvieri*. We then built the first molecular (control region [CR] and rhodopsin [RH]) phylogeny including historical and recently collected representatives of all the *Orestias* complexes as recognized by Parenti (1984a): *agassizii*, *cuvieri*, *gilsoni* and *mulleri*. Our ancient DNA extraction protocol was validated after PCR amplification through an approach based on fragment-by-fragment chimera detection. After optimization, we were able to amplify < 200 bp fragments from both mitochondrial and nuclear DNA (CR and RH, respectively) from probably formalin-fixed type specimens bathed entirely in the extraction fluid. Most of the individuals exhibited few modifications of their external structures after GuSCN bath. Our approach combining type material and 'fresh' specimens allowed us to taxonomically delineate four clades recovered from the well-resolved CR tree into four redefined complexes: *agassizii* (*sensu stricto*, i.e. excluding *luteus*-like species), *luteus*, *cuvieri* and *gilsoni*. The *mulleri* complex is polyphyletic. Our phylogenetic analyses based on both mitochondrial and nuclear DNA revealed a main, deep dichotomy within the genus *Orestias*, separating the *agassizii* complex from a clade grouped under shallow dichotomies as (*luteus*, (*cuvieri*, *gilsoni*)). This 'deep and shallow' diversification pattern could fit within a scenario of ancient divergence between the *agassizii* complex and the rest of *Orestias*, followed by a recent diversification or adaptive radiation within each complex during the Pleistocene, in- and outside the Lake Titicaca. We could not recover the reciprocal monophyly of any of the 15 species or morphotypes that were considered in our analyses, possibly due to incomplete lineage sorting and/or hybridization events. As a consequence, our results starkly question the delineation of a series of diagnostic characters listed in the literature for *Orestias*. Although not included in our phylogenetic analysis, the syntype of *O. jussiei* could not be assigned to the *agassizii* complex as newly defined. The CR sequence of the extinct *O. cuvieri* was recovered within the *cuvieri* clade (same haplotype as one representative of *O. pentlandii*), so the mtDNA of the former species might still be represented in the wild.

Key words: Ancient DNA, ethanol-fixed specimen, formalin-fixed specimen, inter-Andean basin**,** museum collections, phylogeny, pupfishes, species flock, species complex

Accepted by M.R. de Carvalho: 4 Mar. 2013; published: 18 Apr. 2013

Introduction

Specimens stored in museum collections represent a crucial source of morphological and genetic information to tackle evolutionary and taxonomic issues. The added value of museum specimens becomes even more important when considering morphological groups with 'fuzzy' species delimitations (e.g. species complexes), and extinct or very elusive taxa (Chakraborty *et al.* 2006; Stuart *et al.* 2006; Wandeler *et al.* 2007).

Although 'ancient DNA' extraction from museum specimens nowadays is done routinely (Ramakrishnan & Hadly 2009; Shapiro & Hofreiter 2012; Wandeler *et al.* 2007), its level of success heavily depends on the chemical treatments the specimens have been subjected to. The situation is especially complex for fluid-preserved collections, due to the variety of preservatives, methods of fixation and storage conditions that have been used during the last two centuries. It also happens that a specimen or series of specimens are subjected to different preservatives and storage conditions during their time in the collections. Unfortunately, information on the specific chemical treatments applied to collections is rarely available in museums' records (Chakraborty *et al.* 2006; Koshiba *et al.* 1993; Schander & Halanych 2003). The resulting challenge for ancient DNA investigators is to find a balance between the investment of time and resources (i.e., the number of museum specimens to include in the study) and the quality of obtained data (i.e., amplifiable DNA fragments sufficiently informative and taxonomically representative).

Fluid-preserved teleostean collections mostly fall in one of the following categories: ethanol-fixed and -preserved, formalin-fixed and ethanol-preserved, and formalin-fixed and -preserved (Koshiba *et al.* 1993; Raja *et al.* 2011). The use of formalin as a preservative for museum specimens has been a common practice since the late 1800s (De Bruyn *et al.* 2011; Schander & Halanych 2003). DNA extraction and PCR amplification from specimens that were fixed or preserved in formalin may be problematic because of cross-linking between proteins and DNA following DNA fragmentation and nucleotide alteration, often resulting in low amounts and poor quality genomic fragments of small size, varying between 100 and 500 bp (Koshiba *et al.* 1993; Pääbo *et al.* 2004; Schander & Halanych 2003; Wirgin *et al.* 1997). The deleterious effects of formalin on the genome are expected to be higher with a greater time of specimen fixation, making the extraction of DNA unlikely from specimens that were fixed in formalin for long periods of time (Chakraborty *et al.* 2006; Koshiba *et al.* 1993; Raja *et al.* 2011; Zhang 2010).

Many ancient DNA protocols adapted to a variety of museum material (e.g. formalin-fixed and paraffin-embedded tissues, molluscs, and teleostean scales) have been published (e.g. Chase *et al.* 1998; Goelz *et al.* 1985; Impraim *et al.* 1987; Rohland & Hofreiter 2007; Wisely *et al.* 2004; Yue & Orban 2001). However, only few protocols for DNA extraction from fluid-preserved teleostean specimens are available, particularly in the case of formalin-preserved specimens (Chakraborty *et al.* 2006; De Bruyn *et al.* 2011; Raja *et al.* 2011; Shedlock *et al.* 1997; Wirgin *et al.* 1997; Zhang 2010). In addition, most of the methods developed for ancient DNA extraction can be highly invasive, damaging to the specimens or leading to the total destruction of samples, sometimes discouraging museum curators from authorizing such investigations on their collections (Bolnick *et al.* 2012; Hofreiter 2012; Rohland *et al.* 2004; Wisely *et al.* 2004). In the context of fluid-preserved teleostean specimens, non-invasive methods for DNA extraction remain a challenge.

The genus *Orestias* Valenciennes (Cyprinodontidae: Teleostei) comprises pupfishes endemic to the high-latitude lakes and tributary streams from the inter-Andean basin of South America. Its main area of distribution covers the Lake Titicaca and most of the Andean Altiplano, from southern Peru to Bolivia and northeastern Chile. It represents one of the three native teleostean genera endemic to the area (Parenti 1984a; Vila *et al.* 2007; Villwock 1986). *Orestias* pupfishes are externally characterized by the absence of pelvic fins, a reduced and irregular body squamation pattern and a unique head pore pattern (Parenti 1984a; Villwock 1986). Given their great ecological and phenotypic diversity, particularly within the species assemblage from Lake Titicaca, they have been considered as a 'species flock' (Kosswig & Villwock 1964; Villwock 1962; but see Parenti 1894b). The first main revision of *Orestias* was proposed by Tchernavin (1944). The author recognized 20 species structured into four groups, distinguished by overall body shape and/or squamation pattern (Table 1). A later revision from the lesser Lake Titicaca delineated 15 species divided into eight morphological groups on the basis of morphometric and meristic characters (Lauzanne 1982). The most exhaustive taxonomic revision was based on morphometric, osteological and meristic characters and included the complete series of type material (Parenti 1984a). The author delineated 43 species grouped into four monophyletic complexes: *cuvieri* Valenciennes, *mulleri* Valenciennes and *gilsoni* Tchernavin, all endemic to Lake Titicaca and nearby water bodies, and *agassizii* Valenciennes (see Eschmeyer &

TABLE 1. Classification of the species of *Orestias* in the three main taxonomic revisions of the genus. Species described by Valenciennes [in Cuvier and Valenciennes] (1846)[a], Castelnau (1855)[b], Cope (1876)[c], Garman (1895)[d], Eigenmann & Allen (1942)[e], Tchernavin (1944)[f], Arratia (1982)[g], Lauzanne (1982)[h] and Parenti (1984a)[i]. Numbers in parentheses correspond to species groups.

Tchernavin (1944)	Lauzanne (1982)	Parenti (1984a)
		cuvieri complex
cuvieri (1) [a]		*cuvieri*
pentlandii (1) [a]	*pentlandii* (6)	*pentlandii*
	ispi (4) [h]	*ispi*
	forgeti (5) [h]	*forgeti*
		mulleri complex
		gracilis [i]
mulleri (4) [a]	*mulleri* (2)	*mulleri*
crawfordi (4) [f]	*crawfordi* (8)	*crawfordi*
tutini (4) [f]		*tutini*
incae (4) [d]		
		gilsoni complex
gilsoni (3) [f]	*gilsoni* (2)	*gilsoni*
taquiri (3) [f]		*taquiri*
mooni (4) [f]	*mooni* (3)	*mooni*
uruni (3) [f]		*uruni*
minutus (3) [f]		*minutus*
minimus (3) [f]	*minimus* (2)	*minimus*
	tchernavini (2) [h]	*tchernavini*
		tomcooni [i]
		imarpe [i]
		robustus [i]
		agassii complex
agassii (four variants, 2) [a]	*agassii* (7)	*agassii*
		empyraeus [e]
		frontosus [c]
polonorum (2) [f]	*polonorum* (7)	*polonorum*
		elegans [d]
jussiei (one subspecies, 2) [a]	*jussiei* (1)	*jussiei*
		puni [f]
		parinacotensis [g]
		laucaensis [g]
		tschudii [b]
		gymnotus [i]
		hardini [i]
		ctenolepis [i]
		ascotanensis [i]
		richersoni [i]
		multiporis [i]
		mundus [i]
luteus (2) [a]	*luteus* (1)	*ututo* [i]
		silustani [e]
		luteus
albus (2) [a]	*albus* (1)	*rotundipinnis* [i]
	olivaceus (8)	*farfani* [i]
		albus
		olivaceus
Hybrids		Hybrids
cuvieri x *pentlandii*		*cuvieri* x *pentlandii*
affinis [d] = *olivaceus* x *agassii*		
		Nomen nudum
		rospigliossi
		pentlandii var. *fuscus*

Fong 2010 for the correct use of the epithet name), the most speciose (24 species) and widely distributed complex, found in lake Titicaca but also in other hydrological systems within the inter-Andean basin. Since then, three additional species within this complex have been described from Chile (Vila 2006; Vila & Pinto 1986; Vila *et al.* 2011). Taxonomic delimitations within *Orestias* by means of classical taxonomic characters (i.e., morphometrics and meristic counts) have been hotly debated (Lauzanne 1992; Loubens 1989; Lüssen *et al.* 2003; Müller 1993; Villwock 1986; Villwock & Sienknecht 1995; Villwock & Sienknecht 1996), for four main reasons: i) the great morphological variation within the genus, ii) the difficulty in segregating among the different developmental stages (Villwock & Sienknecht 1995; Villwock & Sienknecht 1996), iii) the definition of diagnostic characters from small series of specimens that appear non-reliable when considering larger series (Villwock 1986; Villwock & Sienknecht 1996), and iv) the bias in the calculation of body indexes from warped fluid-preserved specimens (Villwock 1986). In addition, the occurrence of hybridization within and between complexes (Tchernavin 1944; Villwock & Sienknecht 1995; Aspiazu 2002; Esquer Garrigos *et al*. submitted), might blur species boundaries, further complicating the establishment of reliable diagnostic morphological characters. More recently, the phylogenetic systematics of *Orestias* was partially revised through a phylogenetic analysis based on mitochondrial DNA (mtDNA) (Lüssen *et al.* 2003). The authors failed to recover the monophyly of the *agassizii* complex, and showed that the latter should exclude *O. luteus* Valenciennes in order to be phylogenetically valid. Besides, the reciprocal monophyly of several species within the *agassizii* complex was not recovered.

In this context, we aim to evaluate the taxonomic boundaries within the genus *Orestias* in a taxonomically exhaustive (i.e. including representatives from all the complexes), molecular phylogenetic framework. For this, we first apply a modified, non-invasive ancient DNA extraction protocol (Rohland *et al.* 2004) to fluid-preserved specimens representing the type series of *Orestias* kept at the Muséum National d'Histoire Naturelle, Paris, France (MNHN), including the extinct species *O. cuvieri* (Harrison & Stiassny 2004; CREO extinctions database http://creo.amnh.org/pdi.html). This protocol using guanidinium-thiocyanate (GuSCN) was designed for non-destructive DNA extraction from 'hard' material (bones, teeth and skin) of mammalian specimens, but has never been tested on 'soft', fluid-preserved collections. We then build a molecular phylogeny encompassing all the species complexes of *Orestias*, through the combination of both historical (type material) and freshly collected representatives of each complex.

Material and methods

Sample collection. Twenty-eight specimens and 11 species, representing the four traditionally delimited complexes of *Orestias* (*cuvieri*, *mulleri*, *gilsoni*, *agassizii*) were collected from the lake Titicaca and wetlands ('bofedales') from Sur Lipez (Bolivia) between 2008 and 2010 (Table 2). Complexes and species have been identified according to external characters and identification keys given in Parenti (1984a) and Lauzanne (1982), and also by comparison with the teleostean collection of the Limnology Department of the Universidad Mayor de San Andres (UMSA), La Paz, Bolivia. Most of the specimens belonging to the g*ilsoni* and *mulleri* complexes could not be identified to the species level. Such individuals were grouped into 'morphotypes' according to their level of similarity in their external aspect (e.g. body shape, coloration and squamation), and were —whenever possible— related to a species or group of species within the complex to which they belong. Collected specimens were preserved in 99 % ethanol solution and deposited in the collections of the UMSA. A small tissue sample was taken from the dorsal muscle of each specimen for genetic analysis.

Sampling of historical fluid-preserved type specimens stored at MNHN included sixteen individuals representing eleven species as described by Valenciennes (1846) Lauzanne (1981) and Castelnau (1855) (Table 3). No information concerning the specific chemical treatment used for fixation was available. However, the external appearance of some specimens (e.g. with very soft and flaccid bodies; see Table 3) suggested they had been dried out several times and re-hydrated with over-concentrated preservative (Villwock 1986). All the fluid-preserved specimens used in our study were photographed before and after the DNA extraction procedure to visualize potential changes in color, scale loss, degradation or loss of fin rays, resulting from the processing of the individuals. In addition, a 'damage scale' was established to assess dehydration and distortion of body surfaces (1 = no damage, 2 = slight distortion and dehydration of body surface, and 3 = strong dehydration and distortion of body surface; see Table 3).

TABLE 2. List of the specimens used in this study, specifying sample locality and Genbank accession numbers. *Species complexes as defined by Parenti (1984a). Species described by Valenciennes [in Cuvier and Valenciennes] (1846)[a], Castelnau (1855)[b] and Lauzanne (1981)[c]. Numbers in parentheses correspond to the new MNHN catalogue numbers assigned to the specimens that were exposed to GuSCN bath.

Species / Taxon	Sample code	Complex*	Sample locality	Genbank acc. numb.	
				CR	RH
O. agassizii	P234	*agassizii*	Lake Poopo, Bolivia	**KC408006**	**KC408044**
O. agassizii	T113	*agassizii*	Huatajata, Lake Titicaca, Bolivia	**KC408007**	**KC408045**
O. cf polonorum (sensu Lauzanne*)*	H260	*agassizii*	HichuKhota lake, Bolivia	**KC408008**	**KC408046**
O. cf agassizii	KH290	*agassizii*	Khotia, Bolivia	**KC408009**	**KC408047**
O. cf agassizii	K300	*agassizii*	Khara Khota, Bolivia	**KC408010**	**KC408048**
O. cf agassizii	TP323	*agassizii*	Puno Bay, Peru	**KC408011**	**KC408049**
O. cf agassizii	BE1182	*agassizii*	Bofedal Celeste, Bolivia	**KC408012**	**KC408050**
O. cf silustani	1682	*agassizii*	Moho Bay, Peru	**KC408013**	**KC408051**
O. luteus	GENT21	*agassizii*	Huatajata, Lake Titicaca, Bolivia	**KC408014**	**KC408052**
O. luteus	GENT24	*agassizii*	Huatajata, Lake Titicaca, Bolivia	**KC408015**	**KC408053**
O. luteus	T077	*agassizii*	Huatajata, Lake Titicaca, Bolivia	**KC408016**	**KC408054**
O. luteus	U175	*agassizii*	Lake UruUru, Bolivia	**KC408017**	**KC408055**
O. luteus	U217	*agassizii*	Lake UruUru, Bolivia	**KC408018**	**KC408056**
O. albus	1351	*agassizii*	Puno Bay, Peru	**KC408019**	**KC408057**
O. 'gilsoni' morphotype A	1609	*gilsoni*	Huatajata, Lake Titicaca, Bolivia	**KC408026**	**KC408062**
O. 'gilsoni' morphotype A	1612	*gilsoni*	Huatajata, Lake Titicaca, Bolivia	**KC408027**	**KC408063**
O. 'gilsoni' morphotype B	GENT38	*gilsoni*	Huatajata, Lake Titicaca, Bolivia	**KC408024**	**KC408064**
O. 'gilsoni' morphotype B	GENT41	*gilsoni*	Huatajata, Lake Titicaca, Bolivia	**KC408025**	**KC408065**
O. cf crawfordi	1613	*gilsoni*	Huatajata, Lake Titicaca, Bolivia	**KC408028**	**KC408066**
O. ispi	T046	*cuvieri*	Huatajata, Lake Titicaca, Bolivia	**KC408029**	**KC408067**
O. ispi	T047	*cuvieri*	Huatajata, Lake Titicaca, Bolivia	**KC408030**	**KC408068**

.....continued on the next page

TABLE 2. (Continued)

Species / Taxon	Sample code	Complex*	Sample locality	Genbank acc. numb. CR	RH
O. ispi	T054	*cuvieri*	Huatajata lake, Bolivia	**KC408031**	**KC408069**
O. pentlandii	1323	*cuvieri*	Saracocha lake, Peru	**KC408032**	**KC408070**
O. pentlandii	1324	*cuvieri*	Saracocha lake, Peru	**KC408033**	**KC408071**
O. 'mulleri' morphotype A	T093	*mulleri*	Huatajata, Lake Titicaca, Bolivia	**KC408020**	**KC408058**
O. 'mulleri' morphotype A	T094	*mulleri*	Huatajata, Lake Titicaca, Bolivia	**KC408021**	**KC408059**
O. 'mulleri' morphotype B	1408	*mulleri*	Capachica, Lake Titicaca, Peru	**KC408022**	**KC408060**
O. 'mulleri' morphotype B	1412	*mulleri*	Capachica, Lake Titicaca, Peru	**KC408023**	**KC408061**
Syntype of *O. agassizii* [a]	MNHN 1905-0180	*agassizii*	Lake Poopo, Bolivia	**KC408034**	**KC408072**
Syntype of *O. agassizii* [a]	MNHN 1905-0181	*agassizii*	Lake Titicaca, Peru	**KC408035**	**KC408073**
Syntype of *O. agassizii* [a]	MNHN A-9602 (2012-0123)	*agassizii*	Chambira river, Peru	**KC408036**	-
Syntype of *O. tschudii* [b]	MNHN A-9604	*agassizii*	Lake Titicaca, Peru	**KC408037**	-
Syntype of *O. humboldti* (syn. *O. cuvieri*) [a]	MNHN A-9595 (2012-0128)	*cuvieri*	Lake Titicaca, Peru	**KC408038**	-
Syntype of *O. pentlandii* [a]	MNHN 0000-4415	*cuvieri*	Lake Titicaca, Peru	**KC408039**	-
Syntype of *O. pentlandii* [a]	MNHN A-9594	*cuvieri*	Lake Titicaca, Peru	**KC408040**	**KC408074**
Paratype of *O. ispi* [c]	1981-0606 (2012-0127#1)	*cuvieri*	Lake Titicaca, Tiquina, Peru	**KC408041**	-
Paratype of *O. ispi* [c]	1981-0606 (2012-0127#22)	*cuvieri*	Lake Titicaca, Tiquina, Peru	**KC408042**	-
Paratype of *O. forgeti* [c]	1981-0604 (2012-0126#38)	*cuvieri*	Lake Titicaca, Tiquina, Peru	**KC408043**	-

TABLE 3. Success of DNA extraction and PCR amplification in historical *Orestias* specimens (including non-type and types). Species described by Valenciennes [in Cuvier and Valenciennes] (1846)[a], Castelnau (1855)[b] and Lauzanne (1981)[c]. Numbers in parentheses correspond to the new MNHN catalogue numbers assigned to the specimens that were exposed to GuSCN bath. Tissue type: ES - entire specimen, PF - pectoral fin, PFM - pectoral fin including a small fragment of muscle, CP - caudal peduncle, M - muscle. See Material and methods for damage scale.

Specimen	MNHN CG number	Tissue Type	External body appearance before DNA extraction	Sample size	Incubation time	DNA (ng/µl)	External body appearance after DNA extraction (damage scale)	PCR amplification							
								Control region					Rhodopsin		
								1	2	3	4	consensus	1	2	consensus
O. mulleri	1981-1428 # 13	ES	rigid	67.2 mm	20 min	120-120	1	+							
	1981-1428 # 4	ES	rigid	62.1 mm	50 min	83.3-99.7	2	+							
	1981-1428 # 7	ES	rigid	63.5 mm	50 min	19.8-30.5	2	+							
	1981-1428 # 17	ES	rigid	57.1 mm	120 min	1.11-3.73	3	+							
	1981-1428 # 29	ES	rigid	63.9 mm	120 min	12.9-51.6	3	+							
	1981-1428 # 3	CP	rigid	< 15 mm	7 days	0.28	NA	+							
	1981-1428 # 3	PF	rigid	< 15 mm	7 days	0.11	NA	+							
	1981-1428 # 31	CP	rigid	< 15 mm	7 days	0.29	NA	+							
	1981-1428 # 31	PFM	rigid	< 15 mm	7 days	0.2	NA	+							
	1981-1428 # 10	ES	rigid	62.2 mm	10 min	22.4-32.3	1	+							
O. ispi	1622-10-11 # 3	ES	rigid	49.2 mm	10 min	0.19	3	-							
Syntype of *O. agassizii* [a]	1905-0180	ES	rigid	58.1 mm	25 min	63.6-66.6	2	+	+	+	+	validated	+	+	validated
Syntype of *O. agassizii* [a]	1905-0181	ES	rigid	57.1 mm	25 min	81.4-84.2	1	+	+	+	+	validated	+	+	validated
Syntype of *O. agassizii* [a]	A-9602 (2012-0123)	ES	soft and flaccid	60.2 mm	25 min	0.12-0.13	3	+	+	R	+	validated	-	-	NA
Syntype of *O. tschudii* [b]	A-9604	M	soft and flaccid	< 15 mm	7 days	0.24	NA	+	+	+	+	validated	-	-	NA
Syntype of *O. jussiei* [a]	A-9599 (2012-0125)	ES	soft and flaccid	83.1 mm	25 min	0.44 - 0.58	2	R	R	-	+	validated	-	-	NA
Holotype of *O. luteus* [a]	A-9608	M	soft and flaccid	< 15 mm	7 days	0.28-0.30	NA	-	-	-	-	NA	-	-	NA
Syntype of *O. albus* [a]	A-9607	M	soft and flaccid	< 15 mm	7 days	<0.1-0.15	NA	R	R	R	R	chimeric	-	-	NA
Syntype of *O. albus* [a]	A-9607	M	soft and flaccid	< 15 mm	7 days	0.35-0.35	NA	+	R	+	+	chimeric	-	-	NA
Paratype of *O. tchernavini* [h]	1981-0771 (2012-0129#22)	ES	rigid	56.1 mm	25 min	0.17	1	+	R	R	R	chimeric	-	-	NA
Syntype of *O. humboldti* (syn. *O. cuvieri*) [a]	A-9595 (2012-0128)	ES	soft and flaccid	59.3 mm	25 min	<0.1-0.20	3	R	R	R	R	validated	-	-	NA
Syntype of *O. pentlandii* [a]	0000-4415	M	soft and flaccid	< 15 mm	7 days	0.22-0.25	NA	+	+	+	+	validated	-	-	NA
Syntype of *O. pentlandii* [a]	A-9594	M	soft and flaccid	< 15 mm	7 days	0.24	NA	+	+	+	+	validated	+	+	validated
Paratype of *O. ispi* [c]	1981-0606 (2012-0127#1)	ES	rigid	76.9 mm	25 min	0.34-0.4	1	+	R	+	+	validated	-	-	NA
Paratype of *O. ispi* [c]	1981-0606 (2012-0127#2)	ES	rigid	69.1 mm	25 min	0.17-0.85	1	R	R	R	R	validated	-	-	NA
Paratype of *O. forgeti* [c]	1981-0604 (2012-0126#38)	ES	rigid	86.1 mm	25 min	0.78-0.99	1	+	+	R	+	validated	-	-	NA
Syntype of *O. mulleri* [a]	A-9605 (2012-0130)	ES	soft and flaccid	65.5 mm	25 min	0.10-0.13	3	R	-	+	R	chimeric	-	-	NA

[+] Sequence obtained from the 1st PCR amplification; [R] Sequence obtained from PCR re-amplification; No PCR amplification.

DNA extraction and PCR amplification from fresh samples. Genomic DNA was extracted using ABI Prism TM 6100 Nucleic Acid PrepStation (Applied Biosystems, Foster City, CA) following manufacturer's recommendations. We amplified (i) 398 bp of the control region (CR; mitochondrial DNA) using the primer pair L-Smel—H-Smel (Falk *et al.* 2003), and (ii) 689 bp of Rhodopsin (RH; nuclear DNA) using a specific internal primer pair (OrRhF 5'-TTGTCAACCCAGCAGCCTAT-3'–OrRhR 5'-GCCGATGACCATGAGAATG-3') designed from a preliminary alignment of sequences obtained with the primer pair Rh193F –Rh1039R (Chen *et al.* 2003). Transmembrane domains in our amplified sequences, covering domains I–VI, were localized using as a reference the Rhodopsin transcript of *Oryzias latipes* deposited in the Ensembl database (http://www.ensembl.org/index.html; access code: ENSORLT00000013289). PCR amplification were performed in a 25 µl final volume with ~100 ng of template DNA, 0.1 mg/ml BSA, 1X PCR direct loading buffer with $MgCl_2$, 0.25 mM of each dNTP, 0.2 µM each of forward and reverse primers, and 1 U *Taq* DNA Polymerase (Q-BIOgene, Illkirch, France). PCR cycling conditions for CR were as follows: initial denaturation at 94°C for 3 min, followed by 35 denaturation – annealing – extension cycles respectively at 94°C for 30 s, 50 °C for 30 s and 72°C for 30 s, and final extension step at 72°C for 15 min. PCR cycling conditions for RH included an initial denaturation at 94°C for 3 min, followed by 35 denaturation – annealing – extension cycles respectively at 94°C for 45 s, 61°C for 1 min and 72°C for 1 min 15 s, and a final extension step at 72°C for 15 min.

DNA extraction and PCR amplification from historical samples. DNA extraction procedures on historical specimens were conducted in a dedicated ancient DNA box equipped with autonomous ventilation system and UV-irradiation. We adapted the Guanidine thiocyanate (GuSCN)-based, non-invasive ancient DNA protocol of Rohland *et al*. (2004) to fluid-preserved historical specimens, including types. We used seven non-type specimens representing *O. mulleri* and a single individual of *O. ispi* Lauzanne collected by Lauzanne (1981) to optimize the protocol according to three criteria: quantity and quality of extracted DNA, level of damage inflicted to specimens after the extraction bath, and PCR amplification success of a short control fragment of the control region. We varied the initial quantity of tissue used for extraction, from entire specimen to small samples of pectoral fins and caudal peduncle. Samples were gently washed in 1X TE buffer (pH=8) at room temperature during 1 h and immediately after, incubated with soft, flat rotation at 40 °C in 2 ml–50 ml (depending on sample size) of a GuSCN buffer (4M GuSCN, 50 mM Tris pH 8.0, 25 mM NaCl, 1.3% Triton X-100, 2.5 mM PTB, 20 mM EDTA). Negative extraction controls were included for each extraction. Incubation times for the entire specimens varied between 10 and 120 minutes, in 15–50 ml sterile centrifuge tubes with conical bottom (Corning, NY). Small tissue samples were incubated during seven days (Rohland *et al.* 2004). DNA was purified using two successive washes of a chloroform-isoamyl alcohol solution (CIA 96:4), precipitated overnight at 4 °C in isopropanol (2/3 volume), and eluted in 50 µl $_{MQ}H_2O$. The final concentration of extracted DNA was estimated with a Qubit® Fluorometer using the Quant-iTTM dsDNA HS Assay Kit (Invitrogen, Villebon-sur-Yvette, France).

Our first test based on non-type, fluid-preserved historical specimens allowed us to determine the optimal incubation time for entire specimens to 20–25 min (Table 3). We thus proceeded to the DNA extraction of entire type specimens (for type series including more than one representative with total length < 9.5 cm) following the procedure described above. In the case of large type specimens and/or unique holotypes, a sample of muscle from the pleural cavity was dissected for extraction. Extraction baths for entire specimens were generally divided in two volumes after the incubation period, and processed separately. After DNA extraction, specimens were washed in a series of increasingly concentrated baths of ethanol (50%, 60% and 70%) to remove possible traces of GuSCN, before being transferred and stored in ethanol 95%.

PCR amplification of short (< 190 bp) and overlapping CR and RH fragments was done with specific primer pairs (Fig. 1; Table 4) designed from nucleotide sequence alignments obtained from fresh samples, using Primer 3 version 0.4.0 (http://frodo.wi.mit.edu/) and OligoAnalyzer 3.1 (http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/default.aspx). In total, we amplified 398 bp bp for CR and 168 for RH (covering trans-membrane domains I–II) (Fig.1). PCR amplification was performed in a final volume of 25 µl under cycling conditions described above, but containing 1–5 µl of template DNA and 2 U *Taq* DNA Polymerase (Q-BIOgene). Negative and positive controls were included for each series of PCR amplifications. PCR amplifications were done in a separate room to avoid contamination. When PCR products were weak or non-visible, we proceeded to re-amplifications, with the following modifications: [mix] 0.8 µl of PCR product as template, 0.1 µM each of forward and reverse primers, 1 U *Taq* DNA Polymerase (Q-BIOgene), [cycling conditions] 30 denaturation–annealing–extension cycles, and annealing temperatures lowered by 2°C. We assessed congruency among PCR products by

sequencing as many products as possible, obtained from different DNA aliquots, both first round PCRs and/or re-amplifications.

TABLE 4. List of the specific primer pairs designed to amplify the short fragments of the control region (CR) and the rhodopsin (RH) from fluid-preserved specimens of *Orestias*. See also Figure 1.

Locus	Fragment	bp	Primer sequence (5'–3')	Ta (°C)
CR	1	139	L-Smel (Falk *et al*. 2003)	51
			aCR-R1: GTMGGCTYACAYTWCTYTAAATG	
	2	194	aCR-F2: CTAGGRTTCTAAATTAAACYRTTCTTTG	54
			aCR-R2: TAGTAGGGRCATTATMYTKTGATGG	
	3	172	aCR-F3: CCATCAMARKATAATGYCCCTACTA	54
			aCR-R3: CAACCGATGCGATGTTCTTAC	
	4	171	aCR-F4: GCTAAAAACTCATARGTCGAGTTATAC	51
			H-Smel (Falk *et al*. 2003)	
RH	1	184	aRH-F1: TTGTCAACCCAGCAGCCTAT	56
			aRH-R1: GTGAATCCTCCGAACACCAT	
	2	164	aRH-F2: TCTCATTCTTGTCGGCTTCC	51
			aRH-R2: CGTTCATGGAGGTGTACATCG	

Ta: annealing temperature.

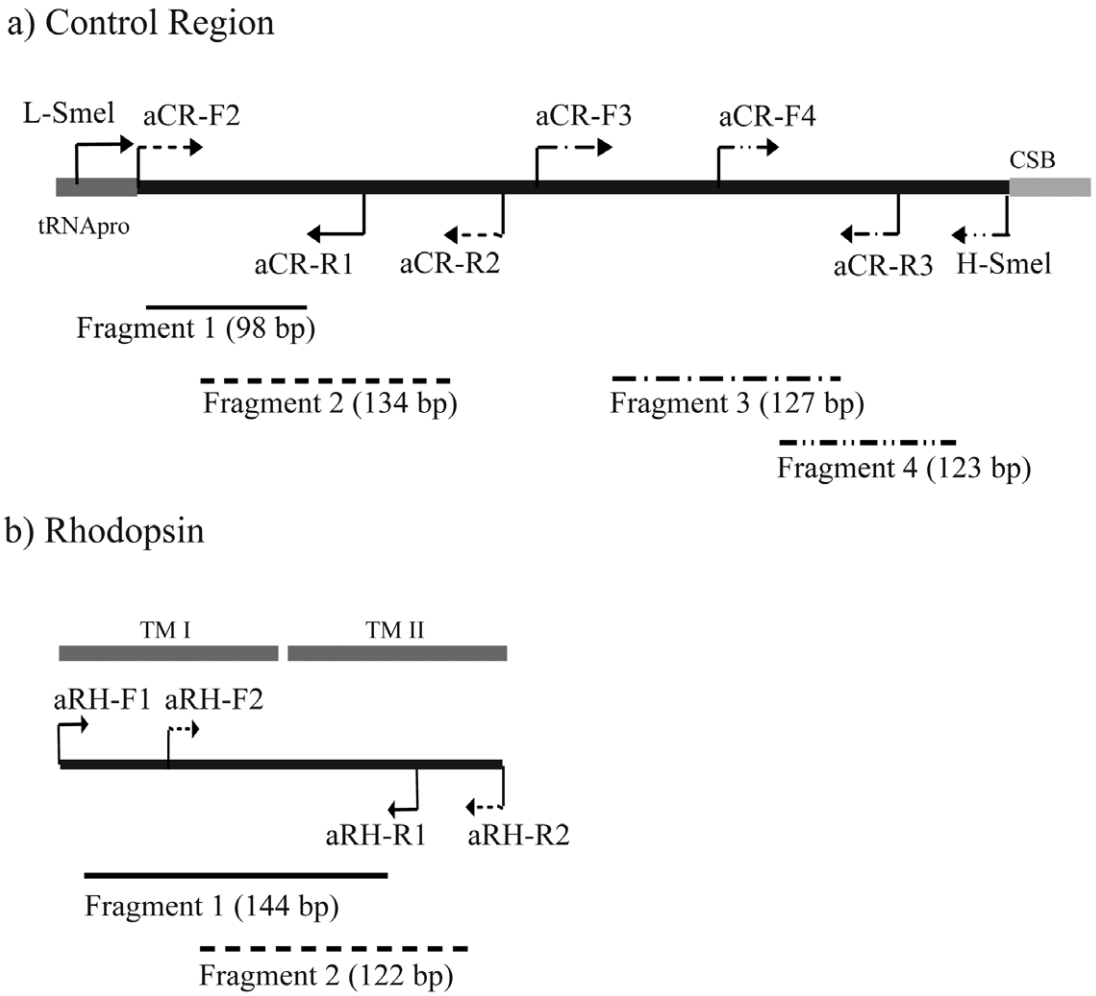


FIGURE 1. Schematic diagram of amplified PCR fragments of control region (a) and rhodopsin (b) from historical specimens, indicating primer location. See Table 4 for primer acronyms.

DNA sequencing and alignment. PCR products were sequenced in both directions on ABI 3730xl DNA Analyzer 96-capillary sequencers (Applied Biosystems) at GENOSCOPE (Consortium National de Recherche en Génomique, Evry, France) and at Eurofins (MWG Operon, Ebersberg, Germany). Nucleotide sequences were edited and aligned by eye using BioEdit 5.0.6 (Hall 1999). Sequences were deposited in Genbank under accession numbers: KC408006 to KC408073 (Table 2).

Validation of the nucleotide sequences obtained from type specimens. In order to identify putative assembly chimeras (composite nucleotide sequences with fragments originating from different organisms) in the nucleotide sequences obtained from type specimens, we ran a pairwise distance analysis on a fragment-by-fragment basis, assessing the match of ancient DNA sequences with the four clades that were recovered in the phylogenetic analysis (see below) using the nucleotide sequence alignment derived from our fresh samples. We used the mean number of pairwise character differences adjusted for missing data (allowing to consider gaps) in PAUP 4.0b10 (Swofford 2003). We returned the set of lowest distances (from 0.0 to 0.03) between each ancient DNA fragment and the series of nucleotide sequences representing the 'fresh' morpho-species. We considered that we obtained a chimeric sequence when at least one of the four fragments had a taxonomic match (criterion of lowest pairwise distance) attributed to a different clade from that of the other fragments.

Whenever chimeric sequences were identified, and if DNA templates were available to allow for additional PCR amplifications, we used the TOPO TA Cloning Kit for Sequencing (Invitrogen) to clone the problematic fragments. Although not detected as chimeric, we also cloned CR fragments obtained from the extinct species *O. cuvieri*. Four to 27 clones per fragment were selected and sequenced using universal M13 primers (Invitrogen).

Phylogenetic reconstruction. Markov Chain Monte Carlo (MCMC) Bayesian inferences were performed to reconstruct the phylogenetic relationships within *Orestias* for both CR and RH sequences, using BEAST 1.6.1 (Drummond & Rambaut 2007). Sequences from type specimens identified as chimeric (on the basis of pairwise distance analysis) or with a high proportion of missing data (e.g. *O. jussiei* Valenciennes; see below) were removed from the analysis.

The most likely evolutionary models were determined in jModelTest 0.1.1 (Posada 2008), using a dataset of the sequences from the 'fresh' specimen without their indels (CR). Model selection was based on the Bayesian Information Criterion (BIC), as performance analysis on simulated datasets has suggested that BIC showed a higher accuracy (Luo *et al.* 2010). The best fitting substitution models were HKY + I for CR and F81 for RH. Single-site gaps in CR alignment were recoded as a different character (base) before running the BEAST analysis. BEAST was run with the following settings: tree prior using the coalescent and assuming constant size (Kingman 1982), and 100,000,000 generations of MCMC steps. Trees and model parameters were sampled every 10,000 generations. Acceptable mixing and convergence to the stationary distribution of the MCMCs was assessed with Tracer 1.5 (Rambaut & Drummond 2007). Two independent runs were performed in order to ensure that posterior probabilities were stable. Log and tree files were concatenated in LogCombiner 1.6.1 (Drummond & Rambaut 2007) with 1,000 generations excluded as burn-in. Trees were summarized as maximum clade credibility trees using TreeAnnotator 1.6.1 (Drummond & Rambaut 2007), and were visualized and edited using Figtree 1.3.1 (Rambaut 2009).

Results

DNA extraction and PCR validation from fluid-preserved specimens

The amount of DNA extracted from entire specimens was significantly higher than for small tissue samples such as caudal peduncle muscle and pectoral fin (one-way ANOVA: F=4.43; P=0.045). However, the amount of extracted DNA was variable in the case of entire specimens, ranging from 0.11 to 120 ng/µl (mean=26.44; Table 3). No correlation was detected among sample size, incubation time and amount of extracted DNA, although this may be due to the overall small number of samples of processed specimens. In general, extracted DNA was highly degraded (< 500 bp; Fig. 2).

Observations of the external appearance of non-type historical specimens of *O. mulleri* showed that a sufficient DNA quantity could be obtained without inflicting significant damage to the external appearance of specimens (but see the case of *O. ispi* MNHN 1622-10-11#3; Table 3), with an incubation time between 20 and 25 min. Longer incubation times inflicted serious damages to the external appearance. More specifically, body surfaces were

highly dehydrated and distorted. The damage levels to the external appearance among type specimens were variable (Fig.3), and ranged from no evident damage (level 1, e.g. *O. forgeti* Lauzanne MNHN2012-0126#38 and *O. ispi* MNHN2012-0127#1 and #2) to very evident dehydration and distortion of the body, mostly for specimens with soft and flaccid bodies (level 3, e.g. *O. humboldti* Valenciennes MNHN 2012-0128). On the other hand, no evident change in colors, scale loss or degradation of fin rays was detected among all the treated historical series.

Irrespective of the extraction conditions (i.e., incubation time, tissue sample and DNA amount) and DNA fragment sizes, most specimens yielded a clear PCR amplification product of the expected size in our first test of amplifying the first fragment of CR (Table 3 and Fig. 2). PCR amplification success rate for the four fragments of CR in type specimens was higher (69 %) than for Rhodopsin (19 %). The CR nucleotide sequences were validated for eleven type specimens, whereas four were identified as chimeric. The consensus sequence obtained for the syntype of *O. jussiei* was considered validated as no conflict between amplified fragments was detected. However, given its level of missing data, this sequence was discarded from the phylogenetic reconstruction (but see Discussion). Within the *cuvieri* complex all the nucleotide sequences obtained from the type specimens were considered validated as no conflict between sequenced fragments was detected. Sequences of the type specimens for *O. tchernavini* Lauzanne (MNHN 1981-0771 // 2012-0129 #22), *O. mulleri* (MNHN A-9605 // 2012-0130), and *O. albus* Valenciennes (MNHN A-9607) were considered chimeric as we detected conflicts among taxonomic attributions of the different fragments constituting their nucleotide sequences (Tables 3 and 5). As a consequence, these specimens were discarded from the phylogenetic reconstruction. The RH sequences that we could obtain for three type specimens representing *agassizii* (*O. agassizii* MNHN 1905-0180 and MNHN 1905-0181) and *cuvieri* complexes (*O. pentlandii* Valenciennes MNHN A-9594) were validated, although taxonomic resolution was lower (Tables 3 and 5; see below). The PCR amplification of the DNA extracted from the holotype of *O. luteus* was not successful.

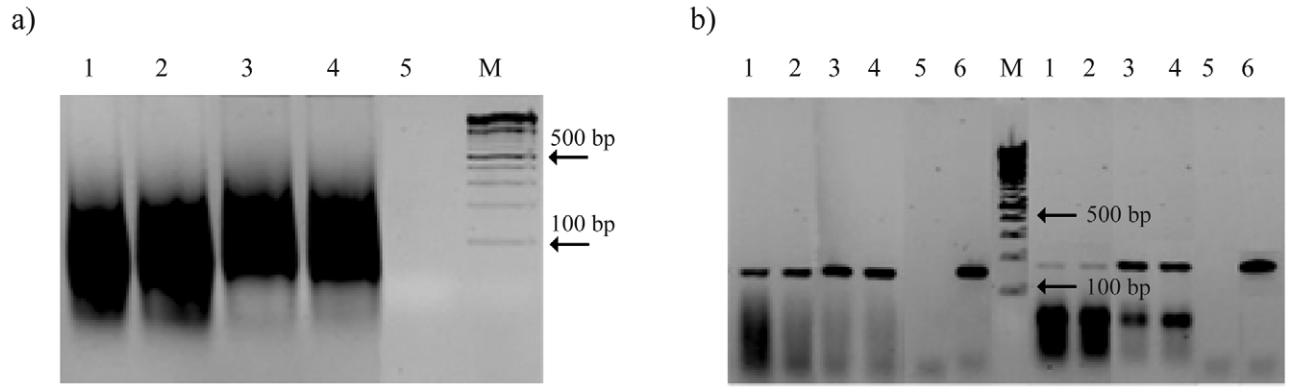


FIGURE 2. Gel electrophoresis displaying extracted DNA (a) and amplified PCR fragments (b) in two historical samples of *O. mulleri* (MNHN 1981-1428#4 and #13). In a, columns 1–2: extracted DNA for specimen #4 (aliquots a and b); columns 3–4: extracted DNA for specimen #13 (aliquot a and b); column 5: negative DNA extraction control; M: molecular weight marker (100 bp ladder). In b, PCR amplification products for control region and rhodopsin are left and right to the molecular weight marker, respectively. Columns 1–2: specimen #4 (aliquots a and b); columns 3–4: specimen #13 (aliquots a and b); column 5: negative PCR control; column 6: positive PCR control. M: molecular weight marker (100 bp ladder).

Phylogeny of the genus *Orestias*

The CR phylogenetic tree recovered four strongly supported clades, branching as follows: Clade A, (Clade B, (Clade C, Clade D))) (Fig.4). We could not recover the reciprocal monophyly of any of the 15 species or morphotypes included in our analysis (Fig. 4). Clade A included all the type specimens and recently collected representatives of *O. agassizii*, *O. cf agassizii* and *O. cf silustani* Allen. Clade B included representatives of *O. luteus*, *O. albus* and two morphotypes attributed to *mulleri* complex. Clade D included the two morphotypes attributed to the *gilsoni* complex and a representative of *O. cf crawfordi* Tchernavin. Clade C included the type specimens of *O. cuvieri*, *O. pentlandii*, *O. ispi* and *O. forgeti*, and all the recently collected representatives of *O. ispi* and *O. pentlandii*. The *mulleri* complex was polyphyletic, with its morphotype representatives spread between clades B and D.

The RH phylogenetic tree was less resolved than the CR tree. Two strongly supported sister-clades were

recovered, Clades 1 and 2 (Fig. 5), corresponding to the mitochondrial Clade A and Clades B-C-D (Fig. 4), respectively. The three type specimens representing *O. agassizii* and *O. forgeti* grouped within the former and latter clades, respectively. The two clades showed a slightly lower support when type specimens were included in the phylogenetic analysis (Fig. 5).

Discussion

Non-invasive ancient DNA protocol applied to fluid-preserved historical specimens of *Orestias*

To our knowledge, our study reports on the first test of non-invasive Guanidine-based protocol (Rohland *et al.* 2004) adapted to fluid-preserved, historical teleostean collection specimen. In our case, i.e. for individuals measuring between 49.2 and 86.1 mm, we suggest that the protocol is 'optimal' (i.e., balance between extracted DNA quantity and damage level of the specimens) when extraction baths of entire specimens are set to 20–25 min. All the historical specimens showed more rigid bodies and appeared 'cleaner' (Rohland *et al.* 2004) after the extraction procedure. Nevertheless, no changes in color, scale loss, and degradation of fin rays were detected, leaving external structures nearly intact for future morphological studies. Our observations are congruent with the generally non-deleterious impact of the Guanidine-based protocol on modern samples of terrestrial arthropods fixed in EtOH 80% (Rowley *et al.* 2007). On the other hand, some *Orestias* type specimens with soft and flaccid bodies seemed more impacted by dehydration and showed severely distorted body surfaces, compared to specimens with originally more rigid bodies (e.g. *O. cuvieri vs. O. forgeti* in Fig.3 and Table 3). According to our observations, variation in the damage level might be attributed to (i) the initial conditions of fixation and storage (which are usually unknown), and (ii) the size of the specimens, since specimens with smaller and thinner bodies were more severely damaged as GuSCN buffer could penetrate the tissues more quickly (e.g. type specimens of *O. cuvieri* 2012-0128 in Fig.3, and specimen of *O. ispi* in Table 3).

FIGURE 3. Example of type specimens treated by incubation in GuSCN bath. Left/right sides: specimens before/after DNA extraction procedure. Damage scale (see Material and Methods): A and B = 1, C = 2 and D = 3. Scale bar represents 2 cm.

Our adaptation of Rolhand *et al.*'s (2004) protocol may prove useful to improve the success of non-invasive DNA extraction from fluid-preserved teleostean specimens. De Bruyn *et al.* (2011) applied an ancient DNA protocol based on silica spin columns (Yang *et al.* 1998) to 80 years old teleostean tissues. This method yielded amplifiable mtDNA fragments from alcohol-fixed specimens, but failed for formalin fixed tissues. The other traditional extraction methods that were also tested (phenol:chloroform, high-salt procedure, QIAmp DNA microkit) proved to be even less efficient. Zhang (2010) reported on a DNA extraction protocol from formalin-preserved teleosteans using a CTAB method with a hot alkali pre-treatment, yielding PCR amplification of fragments of the cytochrome c oxidase subunit 1 (COI). Nevertheless, this method was tested in samples less than 23 years old, and proved to be optimal for samples fixed within the year. Raja *et al.* (2011) reported on the successful DNA extraction from formalin-fixed teleostean tissues using a standard extraction protocol (Nishiguchi *et al.* 2002), but did not discuss the possibility of DNA amplification from the resulting extracts. Our extraction protocol may be relevant for both alcohol and formalin fluid-preserved teleostean specimens. Even if we did not have access to detailed information about the specific fixation conditions of the specimens (a common situation in museum collections; Chakraborty *et al.* 2006; Koshiba *et al.* 1993; Schander & Halanych 2003), the latter showed differences in body stiffness (i.e. flaccid *versus* rigid bodies) suggesting that they were probably subjected to different fixation procedures and storage conditions during their 'life' as vouchers (e.g. ethanol-fixed and -preserved, formalin-fixed and ethanol-preserved, or formalin-fixed and-preserved; see Villwock 1986).

Although the protocol we have set up here can be considered non-invasive, we insist on the fact that the estimation of the optimal incubation time in GuSCN buffer may have to be adjusted depending on (i) the length and width (volume) of the specimen, and (ii) their initial conservation state, e.g. specimens with flaccid body *versus* specimens with more rigid bodies. The protocol was originally designed to extract DNA from 'hard' material of mammalian voucher specimens (bones, teeth and samples of skin; Rohland *et al.* 2004), so we cannot anticipate the potential long-term effects on the conservation of extracted specimens, and the feasibility and advisability of successive DNA extractions from the same specimen. In addition, as GuSCN is a hazardous substance, it is important to take precautions before returning the treated specimens into collections (Bolnick *et al.* 2012), such as washing with diluted ethanol baths, and storing in separate jars to avoid potential diffusion of fluids from the treated specimens.

A total of 11 out of 15 type material nucleotide sequences were validated by our chimera-detection approach (Table 5). Although we consider that our protocol is validated following this procedure, we could not totally remove the incidence of contamination, one of the most difficult aspects when working with ancient or historical samples (Hebsgaard *et al.* 2005; Kelman & Kelman 1999; Yang & Watt 2005). The presence of four chimeric sequences in our alignment could be explained by cross-specimen contamination via organic micro-fragments and mucus through the steps of collective storage during field work and once in the collections (Wandeler *et al.* 2007), and/or the proneness of samples with low DNA quantity and quality, such as historical samples, to exogenous DNA contamination (Pääbo *et al.* 2004) and artifact formation (e.g. chimeras) during PCR amplification. In the latter case, chimeric DNA sequences are produced during "jumping PCR" events, from two or more template molecules owing to incomplete strand synthesis and blocking lesions (Fulton *et al.* 2012; Hebsgaard *et al.* 2005; Kelman & Kelman 1999; Pääbo *et al.* 1989).

In our case, the adaptation of Rohland *et al.*'s (2004) protocol led to the successful amplification of nDNA (RH) fragments. In comparison, Rohland *et al.* (2004) were unable to amplify nDNA from samples of skin, bone and tooth of historical specimens of mammals. The success rate of PCR amplification was clearly higher for mtDNA (CR) than for nDNA (11 *versus* three validated sequences, out of 15 DNA extracts; Table 5). This might be explained by the greater copy number of mtDNA *versus* nDNA (~500-fold; De Bruyn *et al.* 2011). In addition, and independent of DNA quantity, the presence/absence of PCR inhibitors and the level of DNA fragmentation (Fulton *et al.* 2012; Turci *et al.* 2010; Zhang 2010), parameters that were not assessed as part of this study, may have shaped the variation in PCR success rates that we observed among DNA extracts.

Phylogeny of the genus *Orestias*

Our approach combining type material and recently collected specimens allowed us to ascertain the taxonomic delineation of two of the four clades that we recovered from the well-resolved mtDNA tree (Fig. 4). Because Clade A included all the type specimens and recently collected representatives of *O. agassizii*, *O. cf agassizii* and *O. cf silustani*, we assign it to the *agassizii* complex as newly defined, which corresponds to the *agassizii* complex *sensu*

TABLE 5. Minimum pairwise distances calculated between type specimen nucleotide sequences and clade representatives from each fragment of control region (CR) and rhodopsin (RH). *Species complexes as defined by Parenti (1984a). Species described by Valenciennes [in Cuvier and Valenciennes] (1846)[a], Castelnau (1855)[b] and Lauzanne (1981)[c]. Numbers in parentheses correspond to the new MNHN catalogue numbers assigned to the specimens that were exposed to GuSCN bath.

Type specimen	MNHN CG number	Complex *	CR-clade assignment				RH-clade assignment	
			Fragment 1	Fragment 2	Fragment 3	Fragment 4	Fragment 1	Fragment 2
Syntype of *O. agassizii* [a]	1905-0180	*agassizii*	0.000 clade A	0.007 clade A	0.007 clade A	0.000 clade A	0.000 clade A	0.000 clade A
Syntype of *O. agassizii* [a]	1905-0181	*agassizii*	0.000 clade A	0.000 clade A	0.007 clade A	0.000 clade A	0.000 clade A	0.000 clade A
Syntype of *O. agassizii* [a]	A-9602 (2012-0123)	*agassizii*	0.020 clade A	0.007 clade A	0.007 clade A	0.00 clade A		
Syntype of *O. tschudii* [b]	A-9604	*agassizii*	0.000 clade A	0.000 clade A	0.007 clade A	0.000 clade A		
Syntype of *O. jussiei* [a]	A-9599 (2012-0125)	*agassizii*	0.030 clades B, C and D	0.030 clades B, C and D	NA	0.000 clades B, C and D		
Syntype of *O. albus* [a]	A-9607	*agassizii*	0.020 clade D	0.020 clade D and C	0.010 clade A	0.000 clade B		
Syntype of *O. albus* [a]	A-9607	*agassizii*	0.030 clades B and D	0.02 clade D	0.010 clade A	0.008 clade A		
Paratype of *O. tchernavini* [c]	1981-0771 (2012-0129#22)	*gilsoni*	0.010 clade B	0.010 clade C	0.000 clades B, C and D	0.000 clade B		
Syntype of *O. humboldti* (syn. *O. cuvieri*) [a]	A-9595 (2012-0128)	*cuvieri*	0.010 clade C	0.007 clade C	0.000 clade C	0.000 clade C		
Syntype of *O. pentlandii* [a]	0000-4415	*cuvieri*	0.000 clade C	0.000 clade C	0.000 clade C	0.000 clade C		
Syntype of *O. pentlandii* [a]	A-9594	*cuvieri*	0.010 clade C	0.007 clade C	0.000 clade C	0.000 clade C	0.000 clade BCD	0.000 clade BCD
Paratype of *O. ispi* [c]	1981-0606 (2012-0127#1)	*cuvieri*	0.010 clade C	0.007 clade C	0.007 clade C	0.000 clade C		
Paratype of *O. ispi* [c]	1981-0606 (2012-0127#22)	*cuvieri*	0.010 clade C	0.010 clade C	0.000 clade C	0.000 clade C		
Paratype of *O. forgeti* [c]	1981-0604 (2012-0126#38)	*cuvieri*	0.010 clade C	0.010 clade C	0.007 clade C	0.000 clade C		
Syntype of *O. mulleri* [a]	A-9605 (2012-0130)	*mulleri*	0.000 clade B	NA	0.000 clade A	0.000 clade A		

Parenti 1984 minus the *luteus* group. Clade C, including the type specimens of *O. cuvieri*, *O. pentlandii*, *O. ispi* and *O. forgeti*, and all the recently collected representatives of *O. ispi* and *O. pentlandii*, is equivalent to the *cuvieri* complex. Although clades B and D did not include any type material, their taxonomic delineation was also

possible. Clade B included all the representatives of *O. luteus* and *O. albus*, two species that are morphologically clearly distinct from the rest of *Orestias* (Lauzanne 1982), and belong with three other species not analyzed here to the *luteus* group defined by Parenti (1984a) as a monophyletic subdivision within the *agassizii* complex. We therefore delineate Clade B as representing the newly, provisionally defined *luteus* complex. We consider that the presence into this clade of two morphotypes attributed to the *mulleri* complex is evidence for the polyphyly of the *mulleri* complex as defined by Parenti (1984a), since *O. cf crawfordi*, also a representative of the *mulleri* complex, is included within another clade (clade D). Clade D includes the two morphotypes attributed to the *gilsoni* complex, and we consider provisionally that it corresponds to this complex.

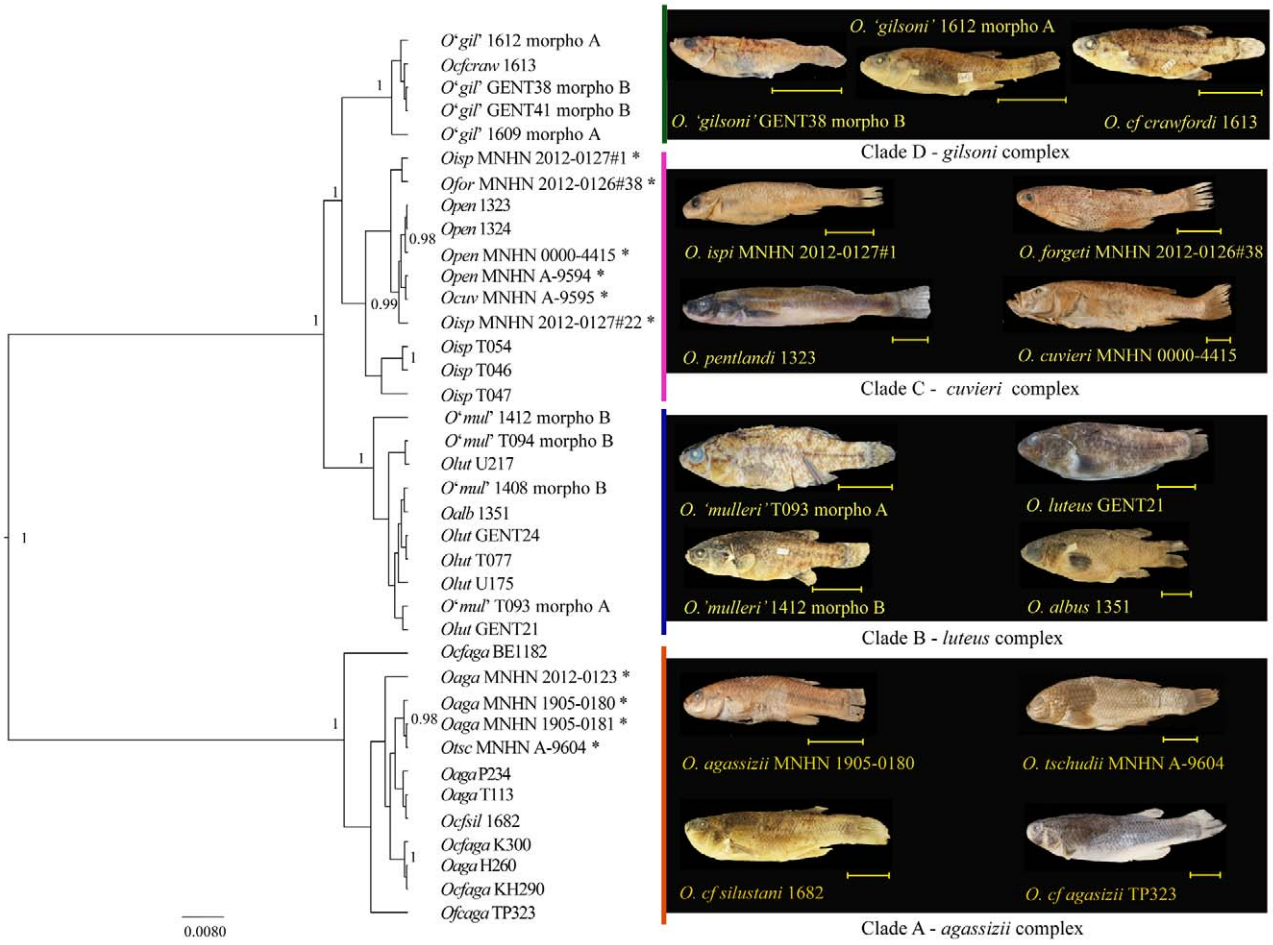


FIGURE 4. Molecular phylogeny of *Orestias* complexes including 'fresh' species and morphotype ("morpho") representatives, and type specimens. Maximum clade credibility tree for control region sequences. Values at branch nodes refer to highest posterior probability of occurrence for clades (> 0.95). Scale bar below tree indicates sequence divergence. Scale bar on pictures represents 2 cm. Type specimens are highlighted with an asterisk. See Table 2 for specimen acronyms. See Discussion for attribution of complex names.

Molecular phylogenetic studies on the genus *Orestias* that would contrast the traditional, morphological-based systematics of the genus are currently lacking or very partial (Lüssen *et al.* 2003; Parker & Kornfield 1995). Our study constitutes the first molecular phylogenetic reconstruction including representatives of all the complexes defined by Parenti (1984a). Our phylogenetic analyses based on both mt- and nDNA revealed a main, deep dichotomy within the genus *Orestias*, separating the *agassizii* complex from a cluster of three clades (Figs. 4 and 5), the latter grouping under shallow dichotomies in the mtDNA tree as complexes (*luteus*, (*cuvieri*, *gilsoni*)). The 'deep and shallow' diversification pattern observed in our mtDNA tree could fit with a scenario of ancient divergence, between the *agassizii* complex and the rest of *Orestias* followed by a recent diversification during the Pleistocene, in- and outside the Lake Titicaca (Lüssen *et al.* 2003; Parenti 1981; Parenti 1984a), possibly as an event of adaptive radiation within each complex (Lauzanne 1982; Lüssen *et al.* 2003; Maldonado *et al.* 2009; Villwock 1986).

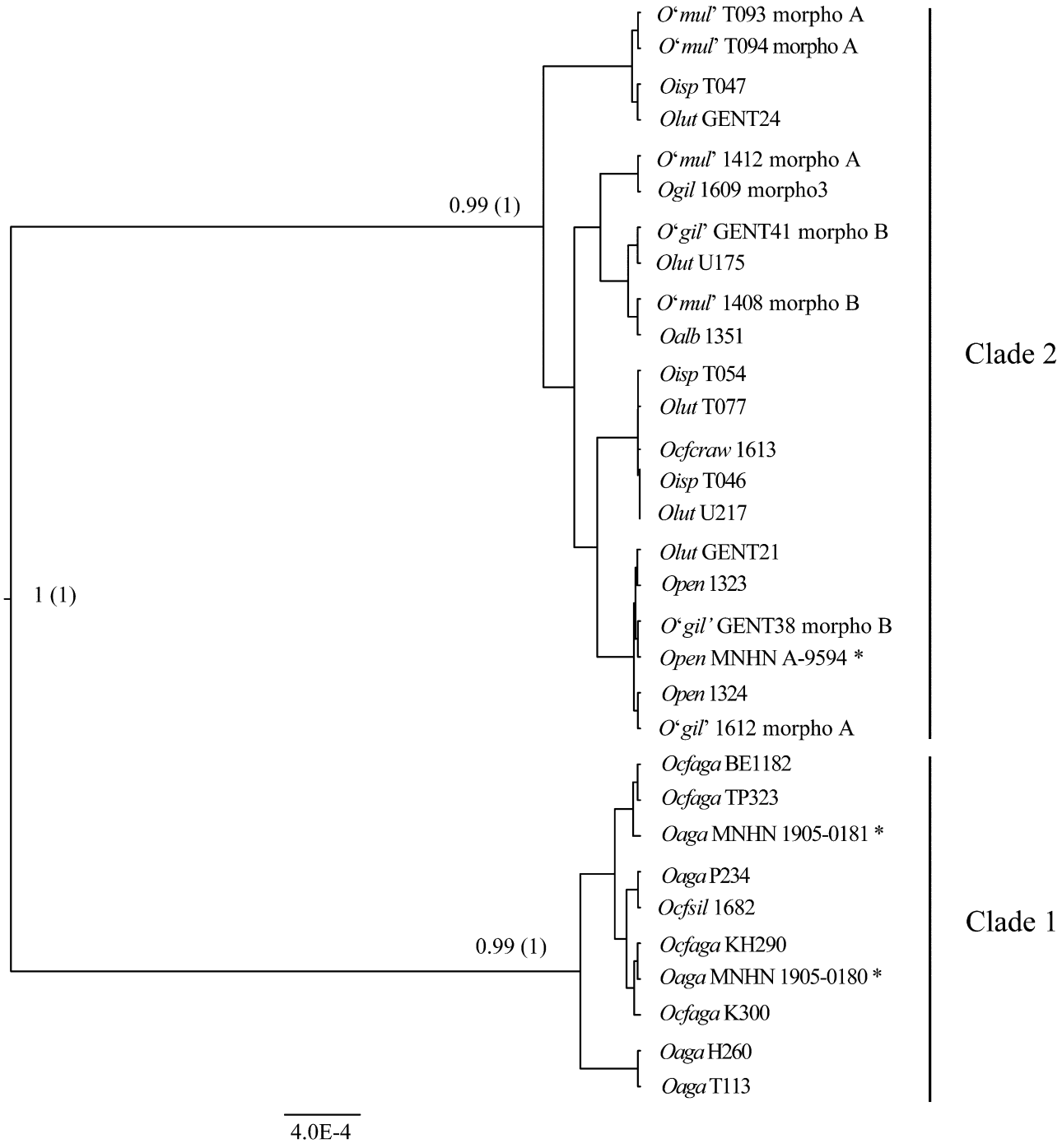


FIGURE 5. Molecular phylogeny of *Orestias* complexes including 'fresh' species and morphotype ("morpho") representatives, and type specimens. Maximum clade credibility tree for rhodopsin sequences. Values at branch nodes refer to highest posterior probability of occurrence for clades (> 0.95); within parenthesis are shown posterior values when removing type specimens from the analysis. Scale bar below tree indicates sequence divergence. Type specimens are highlighted with an asterisk. See Table 2 for specimen acronyms.

We propose, for the first time, a well-supported phylogenetic hypothesis and testing of the monophyly of complexes within the genus *Orestias*. Our results starkly question the use of a series of diagnostic characters established in the literature for complexes and groups of *Orestias* species. As a result, the synapomorphies proposed by Parenti (1984a) for the *agassizii* complex (i.e. including *luteus*-like species), consisting in a series of squamation characters (pattern and scale types) and the presence of a relatively deep caudal peduncle in adults, should be revised given the DNA-based exclusion from the complex of wide head and body morphotypes such as *O. luteus* and *O. albus* (also see Lüssen *et al.* 2003).

The *luteus* complex as defined by our analyses is temporarily left without any diagnostic characters, since it includes a combination of morphotypes (i) with wide head and body, and large and thick granulated scales that partially cover the body (*O. albus* and *O. luteus*), but also (ii) with smaller head and fully scaled body ("*mulleri*" morphotypes A and B). Interestingly, morphotype A shows granulated scales on the head, although these scales are thin and smaller than those present in *O. luteus* and *O. albus*. Parenti (1984a) delimited the *luteus* group from the rest of the species within the *agassizii* complex on the basis of their thick and granular scales present from the snout to the base of the dorsal fin (preopercular, opercular and dorsal region of the pectoral fins). However, granulated scales may also be present on the dorsal surface of the head of adults of *O. cuvieri* and *O. pentlandii* (Parenti 1984a). In addition, in *Orestias* as in other cyprinodontids (e.g. *Aphanius* Nardo), squamation pattern changes with age and growth (Parenti 1984a; Villwock & Sienknecht 1996). Since squamation pattern has been one of the primary characters used to distinguish among *Orestias* species (Parenti 1984a; Tchernavin 1944), its phylogenetic significance should be reconsidered (Villwock and Sienknecth 1996, Müller 1993, this study).

The *cuvieri* complex was fully recovered as a clade by our phylogenetic reconstruction. As a consequence, the five synapomorphies proposed by Parenti (1984a) still stand: i) elongate branchial apparatus with a narrow basihyal, ii) in adults, fifth ceratobranchials narrow and very close together or fused along midline, iii) interarcual cartilage long, thick rod, equal to or longer than first epibranchial bone, iv) interarcual cartilage equal or longer than first epibranchial bone, and v) increase in modal number of anal fin rays (16–17, range: 14 to 19). However, a detailed analysis of these osteological characters on large series of specimens is required to validate their diagnostic values.

The *gilsoni* complex as provisionally defined by our analyses included "*gilsoni*" morphotype B, which morphologically corresponds to a subset of species (*imarpe* Parenti, *robustus* Parenti, *uruni* Tchernavin, *tchernavini* Lauzanne) described by Parenti (1984a) as 'robust' and 'marbled', and morphotype A, which resembles *O. gilsoni*. Parenti (1984a) distinguished *O. gilsoni* from other species within the *gilsoni* complex as having a wide and dorsally flattened head and a narrow dorsal peduncle. The clade attributed to the *gilsoni* complex also included *O. cf crawfordi*, a deep-bodied, laterally compressed, fully scaled species belonging previously to the *mulleri* complex and closely related to *O. incae* Garman (see Parenti 1984a). The *gilsoni* complex was defined by Parenti (1984a) on a single synaphomorphic character state, namely "the procurrent caudal fin rays lying mostly interior to the body profile and closer to the vertebral column, giving the caudal fin a tapered external appearance". From our observations, this character proves very subjective to define and did not represent a reliable diagnostic character. As in the case of the newly defined *luteus* complex, *gilsoni* complex is temporally left without any diagnostic characters.

Eventually, the polyphyly of the *mulleri* complex (Parenti 1984a) as suggested by our analyses seriously questioned the validity of the single synapomorphy on which the complex was based (base of dorsal and anal fins projecting beyond the primary body profile and usually covered with small and thin scales; Parenti 1984a).

Importantly, we could not recover the reciprocal monophyly of any of the morphospecies (15) that were considered in our analyses, despite the fact that we used a fragment of the control region, one of the fastest-evolving mtDNA markers in vertebrates (Avise 2000). The evolutionary picture within the *Orestias* clades (or complexes) may correspond to that observed in recent adaptive radiations (0.001 to 2 Myr) within African Lake cichlids, where the occurrence of incomplete lineage sorting may prevent the recovery of monophyletic species lineages, despite clear-cut morphological, ecological and behavioral delineations among species (e.g. Bezault *et al.* 2011; Meyer *et al.* 1990; Moran & Kornfield 1993; Parker & Kornfield 1997). Moreover, cases of hybridization have been reported, notably between clearly differentiated taxa of *Orestias*. Within the *cuvieri* complex, natural hybrids between the large *O. cuvieri* and *O. pentlandii* were identified by Tchernavin (1944). Between *O. agassizii* and *O. luteus*, two morphologically and ecologically well-differentiated taxa (Maldonado *et al.* 2009), fertile hybrids have been reported based on crossbreeding experiments (Azpiazu 2002) and genetic data (Esquer Garrigos et al. submitted). Tchernavin (1944) and Villwock & Sienknecht (1995) also recorded hybrids *O. olivaceus* Garman x *O. agassizii* and *O. agassizii tschudii* Castelnau x *O. luteus*, respectively. Incidence of hybridization might blur species boundaries, further complicating the establishment of reliable and diagnostic, molecular and morphological characters. Furthermore, hybridization may in turn promote the morphological diversification and partly maintain the disruption of monophyly in *Orestias* morpho-species, notably if complexes are undergoing recent adaptive radiations where interspecific reproductive barriers are not completely efficient (e.g. Herder *et al.* 2006; Shaffer & Thomson 2007). In such a context, the systematics of the genus *Orestias* should be re-evaluated in depth through

more extensive and integrative approaches, including the analysis of co-dominant markers to better estimate gene flow and hybridization within and among complexes, the setting up of breeding experiments in controlled environments, and the search for morphological and osteological diagnostic features to reach a finer characterization of intra and inter-specific variation.

The cases of *O. jussiei* and *O. cuvieri*

Although not included in the phylogenetic analysis, the case of the syntype of *O. jussiei* (MNHN A-9599 // 2012-0125) can be discussed from a taxonomic point of view. Since all the CR fragments amplified from the specimen were assigned to different complexes than *agassizii* (to clades B, C and D at the same time, so without conflict among taxonomic/clade attribution), our results argue for the exclusion of *O. jussiei* from the *agassizii* complex as defined by our phylogenetic analysis. Whereas Parenti (1984a) included *O. jussiei* within the *agassizii* complex (but not within the *luteus* group), Lauzanne (1982) grouped the species with *O. luteus* and *O. albus*. The taxonomic identification of the type specimen of *jussiei* itself seems problematic, since its pronounced dorsal curvature has been suggested to be an artifact following poor fixation procedure (Lauzanne 1982) or a cranial deformation by *Metacercariae* brain infection (Heckmann 1992; Mueller 1972).

One of the most valuable aspects of our work consists in the sequencing of the CR haplotype of the extinct *O. cuvieri*, the most emblematic species within the genus *Orestias*. The nucleotide sequence of *O. cuvieri* was identical to that of the closely related species *O. pentlandii* (MNHN A-9594), and just slightly different (two base pairs) from haplotypes found in type specimens of *O. ispi*, highlighting the very close genetic proximity between morphologically and ecologically well differentiated species within the *cuvieri* complex. One of the probable causes of the extinction of *O. cuvieri* is competitive exclusion or predation through the introduction of non-native species such as trouts, salmonids and silversides between 1940 and 1950 (Lauzanne 1992; Parenti 1984a; Vila *et al.* 2007). However, given the genetic similarity among representatives of the *cuvieri* complex (especially between *O. cuvieri* and *O. pentlandii*), we may consider that the mitochondrial genome of the species is still represented in the wild, whether 'extinction' of the morphotype '*O. cuvieri*' occurred via competitive exclusion (genuine extinction) or hybridization and introgression with congeners ('extinction' by morphotype erosion).

Acknowledgments

We thank Claudia Zepita, Alexander Flores, Kelvin Herbas and Ramón Catari for field assistance (UMSA), Zora Gabsi and Claude Ferrara (Direction des Collections, MNHN) for their kind help in accessing and photographing the *Orestias* collection, and Thierry Oberdorff (UMR BOREA, MNHN) for project management at the early stage. This work was funded by the Agence Nationale de la Recherche (project "FISHLOSS" – ANR-09-PEXT-008) Laboratory work was supported by the 'Consortium National de Recherche en Génomique', and the 'Service de Systématique Moléculaire' of the Muséum National d'Histoire Naturelle (CNRS - UMS 2700). It is part of agreement 2005/67 between Genoscope and the Muséum National d'Histoire Naturelle on the project "Macrophylogeny of life" directed by Guillaume Lecointre, and by the project @SPEED-ID "Accurate SPEcies Delimitation and Identification of eukaryotic biodiversity using DNA markers" proposed by F-Bol, the French Barcode of life initiative.YEG received a PhD-fellowship of the Institut de Recherche pour le Développement-Département Soutien et Formation. We thank Agnès Dettaï for her fruitful comments on the early draft of the manuscript.

References

Arratia, G. (1982) Peces del Altiplano de Chile. *In:* Veloso, A. & Bustos, E. (Eds.) *El Hombre y los Ecosistemas de Montaña. Vol. 1*. Oficina Regional de Ciencia y Tecnologia de la UNESCO para America Latina y el Caribe, Montevideo, Uruguay, pp. 93–133.

Avise, J.C. (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, MA, 447 pp.

Azpiazu, W.I. (2002) *Hibridación artificial entre carachi (Orestias agassii) y punku (Orestias luteus) "killifish" del Lago Titicaca*. Tesis de Licenciatura, Facultad de Agronomia, Universidad Mayor de San Andres, La Paz, Bolivia, 62 pp.

Bezault, E., Mwaiko, S. & Seehausen, O. (2011) Population genomic tests of models of adaptive radiation in Lake Victoria region cichlid fish. *Evolution*, 65, 3381–3397.
http://dx.doi.org/10.1111/j.1558-5646.2011.01417.x

Bolnick, D.A., Bonine, H.M., Mata-Míguez, J., Kemp, B.M., Snow, M.H. & LeBlanc, S.A. (2012) Nondestructive sampling of human skeletal remains yields ancient nuclear and mitochondrial DNA. *American Journal of Physical Anthropology*, 147, 293–300.
http://dx.doi.org/10.1002/ajpa.21647

Castelnau, F. (1855) *Expédition dans les parties centrales de l'Amérique du Sud, de Rio de Janeiro à Lima, et de Lima au Para*. II. Poissons. Paris, 112 pp.

Chakraborty, A., Sakai, M. & Iwatsuki, Y. (2006) Museum fish specimens and molecular taxonomy: a comparative study on DNA extraction protocols and preservation techniques. *Journal of Applied Ichthyology*, 22, 160–166.

Chase, M.R., Etter, R.J., Rex, M.A. & Quattro, J.M. (1998) Extraction and amplification of mitochondrial DNA from formalin-fixed deep-sea molluscs. *BioTechniques*, 24, 243–247.

Chen, W., Bonillo, C. & Lecointre, G. (2003) Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. *Molecular Phylogenetics and Evolution*, 26, 262–288.
http://dx.doi.org/10.1016/S1055-7903(02)00371-8

Cope, E.D. (1876) Note on the ichthyology of Lake Titicaca. *Journal of the Academy of Natural Sciences of Philadelphia*, 8, 185–187.

Cuvier, G. & Valenciennes, A. (1846) *Histoire Naturelle des Poissons*. Tome 18, Paris, 505 pp.

De Bruyn, M., Parenti, L.R. & Carvalho, G.R. (2011) Successful extraction of DNA from archived alcohol-fixed white-eye fish specimens using an ancient DNA protocol. *Journal of Fish Biology*, 78, 2074–2079.
http://dx.doi.org/10.1111/j.1095-8649.2011.02975.x

Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7, 214.

Eigenmann, C.H. & Allen, W.R. (1942) *Fishes of western South America. I. The Intercordilleran and Amazonian lowlands of Peru. II. The high pampas of Peru, Bolivia, and northern Chile. With a revision of the Peruvian Gymnotidae, and of the genus Orestias*. University of Kentucky, 494 pp.

Eschmeyer, W.N. & Fong, J.D. (2010) Species by Family/Subfamily. Online catalogue of fishes. Available from http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp (accessed 25 August 2012)

Falk, T.M., Teugels, G.G., Abban, E.K., Villwock, W. & Renwrantz, L. (2003) Phylogeographic patterns in populations of the black-chinned tilapia complex (Teleostei, Cichlidae) from coastal areas in West Africa: support for the refuge zone theory. *Molecular Phylogenetics and Evolution*, 27, 81–92.
http://dx.doi.org/10.1016/S1055-7903(02)00369-X

Fulton, T.L., Wagner, S.M. & Shapiro, B. (2012) Case study: recovery of ancient nuclear DNA from toe pads of the extinct passenger pigeon. *In:* Shapiro, B. & Hofreiter, M. (Eds.) *Ancient DNA: Methods and protocols*. Humana Press, California, pp. 29–35.
http://dx.doi.org/10.1007/978-1-61779-516-9_4

Garman, S.W. (1895) *The cyprinodonts*. Memoirs of the Museum of Comparative Zoology, Cambridge, 19, pp. 1–179.

Goelz, S.E., Hamilton, S.R. & Vogelstein, B. (1985) Purification of DNA from formaldehyde fixed and paraffin embedded human tissue. *Biochemical and Biophysical Research Communications*, 130, 118–126.
http://dx.doi.org/10.1016/0006-291X(85)90390-0

Hall, T. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95–98.

Harrison, I.J. & Stiassny, M.L.J. (2004) CREO List of Fish Extinctions since AD 1500. American Museum of Natural History, Commitee on Recently Extinct Organism. Available from http://creo.amnh.org/ (accessed 25 August 2012).

Hebsgaard, M.B., Phillips, M.J. & Willerslev, E. (2005) Geologically ancient DNA: fact or artefact? *Trends in Microbiology*, 13, 212–220.
http://dx.doi.org/10.1016/j.tim.2005.03.010

Heckmann, R.A. (1992) Host records and tissue locations for *Diplostomum mordax* (Metacercariae) inhabiting the cranial cavity of fishes from Lake Titicaca, Peru. *Journal of Parasitology*, 78, 541–543.
http://dx.doi.org/10.2307/3283665

Herder, F., Nolte, A.W., Pfaender, J., Schwarzer, J., Hadiaty, R.K. & Schliewen, U.K. (2006) Adaptive radiation and hybridization in Wallace's Dreamponds: evidence from sailfin silversides in the Malili Lakes of Sulawesi. *Proceedings of the Royal Society B: Biological Sciences*, 273, 2209–2217.
http://dx.doi.org/10.1098/rspb.2006.3558

Hofreiter, M. (2012) Nondestructive DNA extraction from museum specimens *In:* Shapiro, B. & Hofreiter, M. (Eds.) *Ancient DNA: methods and protocols*. Humana Press, California, pp. 93–100.
http://dx.doi.org/10.1007/978-1-61779-516-9_13

Impraim, C.C., Saiki, R.K., Erlich, H.A. & Teplitz, R.L. (1987) Analysis of DNA extracted from formalin-fixed, paraffin-embedded tissues by enzymatic amplification and hybridization with sequence-specific oligonucleotides. *Biochemical and Biophysical Research Communications*, 142, 710–716.
http://dx.doi.org/10.1016/0006-291X(87)91472-0

Kelman, L.M. & Kelman, Z. (1999) The use of ancient DNA in paleontological studies. *Journal of Vertebrate Paleontology*, 19, 8–20.
http://dx.doi.org/10.1080/02724634.1999.10011118

Kingman, J.F.C. (1982) The coalescent. *Stochastic Processes and their Applications*, 13, 235–248.
http://dx.doi.org/10.1016/0304-4149(82)90011-4

Koshiba, M., Ogawa, K., Hamazaki, S., Sugiyama, T. & Ogawa, O. (1993) The effect of formalin fixation on DNA and the extraction of high-molecular-weight DNA from fixed and embedded tissues. *Pathology Research and Practice*, 189, 66–72.
http://dx.doi.org/10.1016/S0344-0338(11)80118-4

Kosswig, C. & Villwock, W. (1964) Das problem der intralakustrischen speziation im Titicaca und Lanaosee. *Verhandlungen der Deutschen Zoologischen Gesellschaft*, 28, 95–102.

Lauzanne, L. (1981) Description de trois *Orestias* nouveaux de lac Titicaca, *O. ispi* n. sp., *O. forgeti* n. sp. et. *O. tchernavini* n. sp. (Pisces, Cyprinodontidae). *Cybium*, 5, 71–91.

Lauzanne, L. (1982) Les *Orestias* (Pisces, Cyprinodontidae) du Petit lac Titicaca. *Revue d'Hydrobiologie Tropicale*, 15, 39–70.

Lauzanne, L. (1992) Fish fauna. Native species. The *Orestias*. *In:* Dejoux, C. & Iltis, A. (Eds.) *Lake Titicaca: a synthesis of limnological knowledge*, Kluwer Academic, London, pp. 405–419.
http://dx.doi.org/10.1007/978-94-011-2406-5_10

Loubens, G. (1989) Observations sur les poissons de la partie Bolivienne du lac Titicaca. IV. *Orestias* spp., *Salmo gairdneri* et problemes d'amenagement. *Revue d'Hydrobiologie Tropicale*, 22, 157–177.

Luo, A., Qiao, H., Zhang, Y., Shi, W., Ho, S., Xu, W., Zhang, A. & Zhu, C. (2010) Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. *BMC Evolutionary Biology*, 10, 242.

Lüssen, A., Falk, T. & Villwock, W. (2003) Phylogenetic patterns in populations of Chilean species of the genus *Orestias* (Teleostei: Cyprinodontidae): Results of mitochondrial DNA analysis. *Molecular Phylogenetics and Evolution*, 29, 151–160.
http://dx.doi.org/10.1016/S1055-7903(03)00077-0

Maldonado, E., Hubert, N., Sagnes, P. & De Mérona, B. (2009) Morphology-diet relationships in four killifishes (Teleostei, Cyprinodontidae, *Orestias*) from Lake Titicaca. *Journal of Fish Biology*, 74, 502–520.
http://dx.doi.org/10.1111/j.1095-8649.2008.02140.x

Meyer, A., Kocher, T.D., Basasibwaki, P. & Wilson, A.C. (1990) Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature*, 347, 550–553.
http://dx.doi.org/10.1038/347550a0

Moran, P. & Kornfield, I. (1993) Retention of an ancestral polymorphims in the Mbuna species flock (Teleostei:Cichlidae) of lake Malawi. *Molecular Biology and Evolution*, 10, 1015–1029.

Mueller, J.F. (1972) Cranial deformation in a fish resulting from intracranial parasitism by strigeid metacercariae. *Journal of Parasitology*, 58, 183–184.
http://dx.doi.org/10.2307/3278273

Müller, R. (1993) Critical remarks on the revision of the genus *Orestias* (Pisces Cyprinodontidae) by Parenti (1984). *Zoologische Jahrbücher. Abteilung für Anatomie und Ontogenie der Tiere*, 123, 31–58.

Nishiguchi, M.K., Doukakis, P.E., Egan, M., Kizirian, D., Phillips, A., Prendini, L., Rosenbaum, H.C., Torres, E., Wyner, Y., DeSalle, R. & Giribet, G. (2002) DNA isolation protocols. *In:* DeSalle, R., Wheeler, W. & Giribet, G. (Eds.) *Techniques in molecular systematics and evolution*. Birkhäuser, Germany, pp. 250–287.
http://dx.doi.org/10.1007/978-3-0348-8125-8_12

Pääbo, S., Higuchi, R.G. & Wilson, A.C. (1989) Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. *Journal of Biological Chemistry*, 264, 9709–9712.

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L. & Hofreiter, M. (2004) Genetic analysis from ancient DNA. *Annual Review of Genetics*, 38, 645–679.
http://dx.doi.org/10.1146/annurev.genet.37.110801.143214

Parenti, L.R. (1981) A phylogenetic and biogeographic analysis of cyprinodontiform fishes (Teleostei, Atherinomorpha). *Bulletin of the American Museum of Natural History*, 168, 335–557.

Parenti, L.R. (1984a) A taxonomic revision of the Andean killifish genus *Orestias* (Cyprinodontiformes, Cyprinodontidae). *Bulletin of the American Museum of Natural History*, 178, 107–214.

Parenti, L.R. (1984b) Biogeography of the Andean killifish genus *Orestias* with comments on the species flock concept. *In:* Echelle, A.A. & Kornfield, I. (Eds.) *Evolution of fish species flocks.* University of Maine at Orono Press, pp. 85–92.

Parker, A. & Kornfield, I. (1995) Molecular perspective on evolution and zoogeography of cyprinodontid killifishes (Teleostei; Atherinomorpha). *Copeia*, 1995, 8–21.
http://dx.doi.org/10.2307/1446795

Parker, A. & Kornfield, I. (1997) Evolution of the mitochondrial DNA control region in the mbuna (Cichlidae) species flock of lake Malawi, East Africa. *Journal of Molecular Evolution*, 45, 70–83.
http://dx.doi.org/10.1007/PL00006204
Posada, D. (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, 25, 1253–1256.
http://dx.doi.org/10.1093/molbev/msn083
Raja, M., Muralidharan, M. & Arunachalam, M. (2011) Comparative analysis of DNA extracted from fish species preserved in formalin in two different periods. *Turkish Journal of Biology*, 35, 331–336.
Ramakrishnan, U.M.A. & Hadly, E.A. (2009) Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. *Molecular Ecology*, 18, 1310–1330.
http://dx.doi.org/10.1111/j.1365-294X.2009.04092.x
Rambaut, A. (2009) FigTree. Version 1.3.1. Available from http://tree.bio.ed.ac.uk/software/figtree (accessed May 2012).
Rambaut, A. & Drummond, A.J. (2007) Tracer v.1.5. Computer program and documentation distributed by the authors at http://beast.bio.ed.ac.uk/Tracer (accessed May 2012).
Rohland, N. & Hofreiter, M. (2007) Comparison and optimization of ancient DNA extraction. *BioTechniques*, 42, 343–352.
http://dx.doi.org/10.2144/000112383
Rohland, N., Siedel, H. & Hofreiter, M. (2004) Nondestructive DNA extraction method for mitochondrial DNA analyses of museum specimens. *BioTechniques*, 36, 814 –821.
Rowley, D.L., Coddington, J.A., Gates, M.W., Norrbom, A.L., Ochoa, R.A., Vandenberg, N.J. & Greenstone, M.H. (2007) Vouchering DNA-barcoded specimens: test of a nondestructive extraction protocol for terrestrial arthropods. *Molecular Ecology Notes*, 7, 915–924.
http://dx.doi.org/10.1111/j.1471-8286.2007.01905.x
Schander, C. & Halanych, K.M. (2003) DNA, PCR and formalinized animal tissue - a short review and protocols. *Organisms Diversity and Evolution*, 3, 195–205.
http://dx.doi.org/10.1078/1439-6092-00071
Shaffer, H.B. & Thomson, R.C. (2007) Delimiting species in recent radiations. *Systematic Biology*, 56, 896–906.
http://dx.doi.org/10.1080/10635150701772563
Shapiro, B. & Hofreiter, M. (2012) *Ancient DNA: methods and protocols*. Humana Press, Springer, New York, 247 pp.
Shedlock, A.M., Haygood, M.G., Pietsch, T.W. & Bentzen, P. (1997) Enhanced DNA extraction and PCR amplification of mitochondrial genes from formalin-fixed museum specimens. *BioTechniques*, 22, 394–400.
Stuart, B.L., Dugan, K.A., Allard, M.W. & Kearney, M. (2006) Extraction of nuclear DNA from bone of skeletonized and fluid-preserved museum specimens. *Systematics and Biodiversity*, 4, 133-136.
http://dx.doi.org/10.1017/S1477200005001891
Swofford, D.L. (2003) PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0b10. *Sinauer, Sunderland, MA.*
Tchernavin, V.V. (1944) A revision of the subfamily Orestiinae. *Proceedings of the Zoological Society of London*, 114, 140–233.
http://dx.doi.org/10.1111/j.1096-3642.1944.tb00218.x
Turci, M., Sardaro, M.L.S., Visioli, G., Maestri, E., Marmiroli, M. & Marmiroli, N. (2010) Evaluation of DNA extraction procedures for traceability of various tomato products. *Food Control*, 21, 143–149.
http://dx.doi.org/10.1016/j.foodcont.2009.04.012
Vila, I. (2006) A new species of killifish in the genus *Orestias* (Teleostei: Cyprinodontidae) from the Southern High Andes, Chile. *Copeia*, 472–477.
http://dx.doi.org/10.1643/0045-8511(2006)2006[472:ANSOKI]2.0.CO;2
Vila, I., Pardo, R. & Scott, S. (2007) Freshwater fishes of the Altiplano. *Aquatic Ecosystem Health and Management Society*, 10, 201–211.
http://dx.doi.org/10.1080/14634980701351395
Vila, I. & Pinto, M. (1986) A new species of Killifish (Pisces, Cyprinodontidae) from the Chilean Altiplano. *Revue d'Hydrobiologie Tropicale*, 19, 233–239.
Vila, I., Scott, S., Mendez, M.A., Valenzuela, F., Iturra, P. & Poulin, E. (2011) *Orestias gloriae*, a new species of cyprinodontid fish from saltpan spring of the southern high Andes (Teleostei: Cyprinodontidae). *Ichthyological Exploration of Freshwaters*, 22, 345–353.
Villwock, W. (1962) Die Gattung *Orestias* (Pisces, Microcyprini) und die Frage der intralakustrischen Speziation im Titicaca-Seengebiet. *Verhandlungen der Deutschen Zoologischen Gesellschaft Wien 1962, Zoologischer Anzeiger*, 26, 610–624.
Villwock, W. (1986) Speciation and adaptive radiation in the Andean *Orestias* fishes. *In:* Vuilleumier, F. & Monasterio, M. (Eds.) *High altitude tropical biogeography.* Oxford University Press, Oxford, pp. 387–403.
Villwock, W. & Sienknecht, U. (1995) Intraspezifische variabilität im genus *Orestias* Valenciennes, 1839 (Teleostei: Cyprinodontidae) und zum problem der artidentität. *Mitteilungen aus dem Hamburgischen Zoologischen Museum und Institut*, 92, 381–398.
Villwock, W. & Sienknecht, U. (1996) Contribucion al conocimiento e historia de los peces Chilenos. Los Cyprinodontidos del género *Orestias* Val. 1839 (Teleostei: Cyprinodontidae) del Altiplano Chileno. *Medio Ambiente*, 13, 119–126.

Wandeler, P., Hoeck, P.E.A. & Keller, L.F. (2007) Back to the future: museum specimens in population genetics. *Trends in Ecology and Evolution*, 22, 634–642.
http://dx.doi.org/10.1016/j.tree.2007.08.017

Wirgin, I., Maceda, L., Stabile, J. & Mesing, C. (1997) An evaluation of introgression of Atlantic coast striped bass mitochondrial DNA in a Gulf of Mexico population using formalin-preserved museum collections. *Molecular Ecology*, 6, 907–916.
http://dx.doi.org/10.1046/j.1365-294X.1997.00271.x

Wisely, S.M., Maldonado, J.E. & Fleische, R.C. (2004) A technique for sampling ancient DNA that minimizes damage to museum specimens. *Conservation Genetics*, 5, 105–107.
http://dx.doi.org/10.1023/B:COGE.0000014061.04963.da

Yang, D.Y., Eng, B., Waye, J.S., Dudar, J.C. & Saunders, S.R. (1998) Improved DNA extraction from ancient bones using silica-based spin columns. *American Journal of Physical Anthropology*, 105, 539–543.
http://dx.doi.org/10.1002/(SICI)1096-8644(199804)105:4<539::AID-AJPA10>3.0.CO;2-1

Yang, D.Y. & Watt, K. (2005) Contamination controls when preparing archaeological remains for ancient DNA analysis. *Journal of Archaeological Science*, 32, 331–336.
http://dx.doi.org/10.1016/j.jas.2004.09.008

Yue, G.H. & Orban, L. (2001) Rapid Isolation of DNA from Fresh and Preserved Fish Scales for Polymerase Chain Reaction. *Marine Biotechnology*, 3, 199–204.
http://dx.doi.org/10.1007/s10126-001-0010-9

Zhang, J. (2010) Exploiting formalin-preserved fish specimens for resources of DNA barcoding. *Molecular Ecology Resources*, 10, 935–941.
http://dx.doi.org/10.1111/j.1755-0998.2010.2838.x

II.3 Article 2

Yareli Esquer Garrigos, Josie Lambourdière, Carla Ibañez & Philippe Gaubert. (2011) **Characterization of ten polymorphic microsatellite loci in the Andean pupfish *Orestias agassizii*, with cross-amplification in the sympatric *O. luteus***. Conservation Genetics Resources, 3, 117-119.

Conservation Genet Resour (2011) 3:17–19
DOI 10.1007/s12686-010-9269-3

TECHNICAL NOTE

Characterization of ten polymorphic microsatellite loci in the Andean pupfish *Orestias agassizii*, with cross-amplification in the sympatric *O. luteus*

Yareli Esquer Garrigos · Josie Lambourdière · Carla Ibañez · Philippe Gaubert

Received: 16 June 2010 / Accepted: 22 June 2010 / Published online: 4 July 2010

Abstract We describe the isolation and characterization of ten polymorphic microsatellite loci in the Andean pupfish *Orestias agassizii*. The number of alleles per locus in 25 individuals was high, varying from 5 to 25. Observed and expected mean heterozygosities were 0.750 and 0.767, respectively. The ten loci conformed to the Hardy–Weinberg equilibrium, and we did not detect linkage disequilibrium between loci. Cross-species amplification in *O. luteus* was entirely successful, and polymorphism was detected in all loci. One locus (A116) showed significant departure from Hardy–Weinberg equilibrium due to homozygote excess. These microsatellites will be useful for assessing the genetic structure and historical demography of Andean pupfishes, which are subject to a global collapse of their stocks.

Keywords Cyprinodontidae · *Orestias agassizii* · Andean altiplano · Microsatellites · Heterozygosity · Cross-amplification · *Orestias luteus*

Y. Esquer Garrigos (✉) · P. Gaubert
UMR BOREA IRD 207, Muséum National d'Histoire Naturelle, 43 rue Cuvier, 75005 Paris, France
e-mail: esquer@mnhn.fr

J. Lambourdière
Département Systématique et Evolution, SSM-UMS 2700, Muséum National d'Histoire Naturelle, 43 rue Cuvier, 75005 Paris, France

C. Ibañez
Institut de Recherche pour le Développement, Instituto de Ecología, Unidad de Limnología, Universidad Mayor de San Andrés, Casilla, 10077 La Paz, Bolivia

The Andean pupfishes *Orestias* spp. (Cyprinodontidae) are endemic to the inter-Andean basin in South America, and constitute an exceptional case of vertebrate radiation at high altitude (Villwock 1986; Maldonado et al. 2009). From the mid-XXth century, the motorization of fishing techniques combined to the introduction of invasive species (silversides and trouts), growing water demand, the eutrophication of lakes and water pollution have resulted in the diminution of *Orestias* stocks (Vila et al. 2007). *Orestias agassizii*—see Eschmeyer and Fong (2010) for the correct use of the specific name—is the most widely distributed *Orestias* 'species complex' (Parenti 1984) and as such, is a traditional, important component of Andean fisheries (Capriles et al. 2008). Although knowledge on the biology and ecology of *Orestias* species remains very fragmentary, it appears that *O. agassizii* fits a trend of global stock decrease throughout its range (Vila et al. 2007).

Microsatellite libraries were developed by Genetic Identification Services (GIS; http://www.genetic-id-services.com) following Jones et al. (2002). Genomic DNA was obtained from one individual of *O. agassizii* identified after Parenti (1984); approximately 100 μg of DNA was digested using a cocktail of seven blunt-end cutting enzymes. Four libraries were enriched for -CA, -GA, -CATC and -TAGA microsatellite repeats. A total of 137 non-redundant clones ranging from 350 to 700 pb were sequenced on a ABI 377 DNA Sequencer (Applied Biosystems, Carlsbad, CA). PCR primer pairs were designed for 105 microsatellite-containing sequences using DesignerPCR version 1.03 (Research Genetics, Huntsville, AL), and a preliminary, visual polymorphism assessment of 24 loci was provided.

We amplified a first set of 33 loci via PCR gradient on three individuals originating from three different lakes (Titicaca, Poopo and Uru Uru, Bolivia). A final subset of ten microsatellite loci showing efficient PCR amplification

Table 1 Characteristics of the ten microsatellite loci isolated from *Orestias agassizii*, and cross-amplified in *O. luteus*

Locus/ Genbank accession no.	Repeat motif	Primer sequence (5′–3′)	Dye label	Ta (°C)	Allele size range (bp) *O. agassizii* *O. luteus*	N_A (*n* individuals) *O. agassizii* *O. luteus*	H_O *O. agassizii* *O. luteus*	H_E *O. agassizii* *O. luteus*	HWE *P*-value *O. agassizii* *O. luteus*
A9a	$CA_5AA_1CA_{11}$	F: CAGGAAGGAATCTCAGGAATG	FAM	58	244–266	11 (24)	0.833	0.848	0.152
HM486404		R: ACGCACCGTTTCATAGTAAGG			234–260	4 (19)	0.158	0.154	1.000
A106	CA_{68}	F: TGGCTGATGGTATTGGTTG	VIC	60	256–317	23 (24)	0.917	0.944	0.794
HM486405		R: AGCACACCTTCACAGGATG[a]			263–313	18 (17)	1.000	0.954	1.000
A116	$CA_5GC_1CA_6CC_1CA_{14}$	F: TCGCTACTTACTCCGACCTC	PET	54	300–392	25 (25)	0.958	0.951	0.056
HM486406		R: AAATCACAATGGCTTTCTCTG			288–334	11 (21)	0.762	0.823	0.000*
B1	$CT_2TT_1CT_{14}$	F: TACAAACACATCCATCTCAGTC	PET	58	171–181	6 (25)	0.560	0.516	0.780
HM486407		R: AACACTCCTATCATCCATCATC			171–177	5 (21)	0.429	0.476	0.835
B103	$CT_3CCATCG_1CT_{14}$	F: TATTATCCACTCCTGGTCAGTC	FAM	51	193–215	6 (25)	0.600	0.610	0.624
HM486408		R: GTTGAAGCGTTTCCAAGAT			195–217	5 (21)	0.810	0.677	0.529
B104	$CT_2CG_1CT_4...CT_{10}$	F: ACCGTAGTTGCCTGGTTACA	VIC	64	141–169	11 (25)	0.920	0.808	0.223
HM486409		R: AGGGTGCTGTCAGAGATGAG			141–157	2 (20)	0.050	0.050	–
C101	$CCAT_{25}CCAC_2CCAT_2$	F: ATTTCAGCAGAGTTTTTGGTAC	NED	57	214–306	17 (25)	0.800	0.922	0.220
HM486410		R: AGCTTCTACTCAGACAAGAGGA			254–318	15 (20)	0.931	0.900	0.094
C102	$CCAT_{15}$	F: TTCCAAACCACATTTTAGATCC	NED	63	225–273	5 (25)	0.200	0.291	0.046
HM486411		R: CAGCCTTTTGATTATGGAGGT			233–281	11 (21)	0.952	0.900	0.174
C105	$CCAT_1CCAA_1CCAT_8$	F: AGCAAGACCAGTTTGAAATCT	NED	58	192–262	12 (25)	0.840	0.838	0.794
HM486412		R: GTTGCCCTGCGATGTAC			200–248	14 (20)	0.850	0.929	0.357
D110	$GATA_{26}$	F: ATCACAAGACGAGGTTCTCAC	FAM	59	232–350	20 (25)	0.880	0.944	0.294
HM486413		R: GATTGGAGCAAGGGACTG			244–314	15 (21)	0.952	0.915	0.900

* Locus showing significant departure from HWE

[a] Pigtail sequence -GTTTCTT added at 5′

were analyzed in 25 individuals of *O. agassizii* from lake Uru Uru (S 18°07′, W 67°05′), using 5′-fluorescently labeled forward primers. Amplification reactions were carried out in a 10 µl final volume, containing ~50 ng of template DNA, 1X PCR buffer (20 mM Tris–HCl pH 8.4, 50 mM·KCl), 2 mM·$MgCl_2$, 0.1 mg/ml BSA, 0.25 mM of each dNTP, 0.2 µM each of forward and reverse primers, and 0.5 U *Taq* DNA Polymerase (Q-BIOgene, Illkirch, France). The PCR cycling conditions were as follows: initial denaturation at 94°C for 2 min, followed by 35 cycles at 92°C for 30 s, 30 s at locus-specific annealing temperature (see Table 1) and extension at 72°C for 1 min, and a final extension step of 30 min at 72°C.

PCR products were analyzed by electrophoresis on an ABI PRISM 3100 Genetic Analyzer using internal size standard GeneScan500 LIZ (Applied Biosystems), and scored using GENEMAPPER 3.0 (Applied Biosystems) and FLEXIBIN 2.0 (Amos et al. 2007). Allele number, observed (H_O) and expected (H_E) heterozygosities were calculated under CERVUS 3.0.3 (Marshall et al. 1998). We estimated deviations from Hardy–Weinberg equilibrium (HWE) and genotypic linkage disequilibrium (LD) under GENEPOP 1.2 (Raymond and Rousset 1995). *P* values were adjusted for multiple test of significance using the sequential Bonferroni correction at the 5% nominal level (Rice 1989). Incidence of null alleles and scoring errors due to stuttering or large allele drop-out were assessed on MICRO-CHECKER 2.2.3 (Oosterhout et al. 2004).

Microsatellite loci isolated from *O. agassizii* showed medium to high levels of allelic diversity. The number of alleles ranged from 5 to 25 (mean = 13.60), and the observed and expected heterozygosities ranged from 0.200 to 0.958 (mean = 0.750) and from 0.291 to 0.951 (mean = 0.767), respectively. We did not detect any deviation from Hardy–Weinberg equilibrium or linkage disequilibrium between pairs of loci after sequential Bonferroni correction ($P > 0.005$; Table 1). The expected distribution of homozygote size classes calculated with MICRO-CHECKER suggested no incidence of null alleles or scoring errors due to large allele drop-out or stuttering.

Using the same PCR conditions as described above, cross-amplification in the sympatric *O. luteus* ($n = 21$) was successful and evidenced specific polymorphism in the ten loci. Mean number of alleles was 10, and mean expected and observed heterozygosities were 0.6778 and 0.689, respectively. Locus A116 significantly deviated from Hardy–Weinberg equilibrium (Table 1) because of homozygote excess. No linkage disequilibrium between pairs of loci and no incidence of null alleles were detected. The ten microsatellite loci that we identified will be useful to assess the genetic structure and demographic history of *O. agassizii* and *O. luteus*. Those traditional components of Andean fisheries are subject to a continuous, global stock decrease. Microsatellite-based studies should contribute to better characterize their population statuses and to help formulate adequate conservation measures.

Acknowledgments We thank Thierry Oberdorff and Bernard Hugueny for management of the *Orestias* project. We are thankful to Claudia Zepita, Alexander Flores and Ramón Catari for field assistance. Mark Todd gave useful comments on an early draft of the manuscript. This work was funded by the program "Action incitative 2008: Diversité des poissons—n°45/Radiation adaptative des *Orestias* dans les lacs andins", from the Département Ressources Vivantes—Institut de Recherche pour le Développement, and Agence Nationale de la Recherche (project "FISHLOSS"—ANR-09-PEXT-008-01). This publication was made possible through support provided by the Institut de Recherche pour le Développement—Département Soutien et Formation to YEG.

References

Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007) Automated binning of microsatellite alleles: problems and solutions. Mol Ecol Notes 7:10–14

Capriles JM, Domic AI, Moore KM (2008) Fish remains from the Formative Period (1000 BC–AD 400) of Lake Titicaca, Bolivia: zooarchaeology and taphonomy. Quatern Int 180:115–126

Eschmeyer W, Fong J (2010) Species of fishes by family/subfamily. Online catalogue of fishes. http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp. Accessed 02 June 2010

Jones KC, Levine KF, Banks JD (2002) Characterization of 11 polymorphic tetranucleotide microsatellites for forensic applications in California elk (*Cervus elaphus canadensis*). Mol Ecol Notes 2:425–427

Maldonado E, Hubert N, Sagnes P, De Mérona B (2009) Morphology-diet relationships in four killifishes (Teleostei, Cyprinodontidae, *Orestias*) from Lake Titicaca. J Fish Biol 74:502–520

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639–655

Oosterhout C, Hutchinson W, Wills D, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538

Parenti LR (1984) A taxonomic revision of the Andean killifish genus *Orestias* (Cyprinodontiformes, Cyprinodontidae). Bull Am Mus Nat Hist 178:107–214

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43:223–225

Vila I, Pardo R, Scott S (2007) Freshwater fishes of the Altiplano. Aquat Ecosyst Health Manag 10:201–211

Villwock W (1986) Speciation and adaptive radiation in the Andean *Orestias* fishes. In: Vuilleumier F, Monasterio M (eds) High altitude tropical biogeography. Oxford University Press, Oxford, pp 387–403

II.4 Article 3

Yareli Esquer Garrigos, Bernard Hugueny, Carla Ibañez, Kellie Koerner, Claudia Zepita, Josie Lambourdiere, Arnaud Couloux & Philippe Gaubert. **Hybridization within an adaptive radiation framework: evolutionary and conservation implications for two species of Andean pupfishes (*Orestias* spp., Teleostei, Cyprinodontidae)**. Molecular Ecology (submitted).

Hybridization within an adaptive radiation framework: evolutionary and conservation implications for two species of Andean pupfishes (*Orestias* spp., Teleostei, Cyprinodontidae)

YARELI ESQUER GARRIGOS,[1*] BERNARD HUGUENY,[1] CARLA IBAÑEZ,[2], KELLIE KOERNER,[3,1] CLAUDIA ZEPITA,[2] JOSIE LAMBOURDIÈRE,[4] ARNAUD COULOUX[5] & PHILIPPE GAUBERT[1]

[1]*UMR BOREA, Département Milieux et Peuplements Aquatiques, MNHN-CNRS 7208-IRD 207-UPMC, Muséum National d'Histoire Naturelle, 43 rue Cuvier, 75231 Paris, France,* [2]*Instituto de Ecología-Unidad de Limnología, Universidad Mayor de San Andrés, Casilla 10077, La Paz, Bolivia,* [3]*Current address: Charter School of the Dunes, 860 N Lake Street, Gary, IN 46403, USA,* [4]*Département Systématique et Evolution, SSM-IFR 101, Muséum National d'Histoire Naturelle, 43 rue Cuvier, 75005 Paris, France,* [5]*Genoscope, Centre National de Séquençage, 2 rue Gaston Crémieux, CP5706, 91057 Evry Cedex, France*

*Correspondence: Yareli Esquer Garrigos, Fax: +33 (0) 1 40 79 37 71; E-mail: esquer@mnhn.fr

Keywords: adaptive radiation, conservation, fitness, hybridization, morphology, *Orestias*.

Running Title: Hybridization in *Orestias*

Abstract

Because hybridization may promote evolutionary novelties but also species extinction, the characterization of hybrids has major implications in evolutionary and conservation issues. Our study focused on two species of Andean pupfishes (*Orestias agassizii* and *O. luteus*), which represent a unique case of adaptive radiation in high-altitude freshwaters. The two species are eco-morphologically well differentiated but yield viable hybrids in experimental conditions. We combined mitochondrial and nuclear sequences, microsatellites, morphological characters and morphometrics to characterize the incidence of natural hybridization between *O. agassizii* and *O. luteus*. We discussed the potential role of hybridization on the adaptive diversification of *Orestias*. We found a high incidence (c. 10%) of natural hybridization between the two species. The combined use of heterozygosity patterns from nuclear sequences (rhodopsin) and assignment procedures using 10 microsatellite loci proved an efficient genetic diagnostic 'tool' for detecting F1 hybrids. Because hybrids deviated from morphological intermediacy (closer to *O. agassizii*), the morphological criterion appears inapplicable to detect *Orestias* hybrids. Our study and experimental evidence suggest weak reproductive isolation barriers within the genus *Orestias*, fitting the evolutionary framework of an ongoing adaptive radiation. We detected a higher proportion of male *O. agassizii* x female *O. luteus* hybrids, suggesting a 'hybrid vigor' effect in natural conditions (in line with reported experimental observations). The morphological proximity of hybrids with *O. agassizii* may reflect the incidence of natural selection in favor of the most ecologically successful *Orestias* morphotype. In terms of conservation, our results provide useful tools for the ongoing restocking programs of *Orestias*.

Keywords: adaptive radiation, conservation, fitness, hybridization, morphology, *Orestias*.

Introduction

The role of hybridization —here defined as the interbreeding of individuals from distinct populations regardless of their taxonomic status (Rhymer & Simberloff 1996) — in speciation and morphological diversification has been a central question in evolutionary biology (Arnold 1992; Seehausen 2004; Mallet 2007; Schwenk *et al.* 2008; see Abbott *et al*. 2013 for a review). Although potentially associated to deleterious effects such as reduced fertility and fitness, loss of genetic integrity and homogenization of gene pools (Rhymer & Simberloff 1996; Seehausen 2006), hybridization has been evidenced as a powerful mechanism of species diversification capable of generating phenotypic novelty and new genotypic variants (Dowling & Secor 1997; Seehausen 2004; Bell & Travis 2005; Schwenk *et al.* 2008; Hayden *et al.* 2010; Tobler & Carson 2010; Stemshorn *et al.* 2011).

Natural hybridization in freshwater teleosteans is common (Hubbs 1955; Scribner *et al.* 2001) and seems more frequent than in any other groups of vertebrates (Schwenk *et al*. 2008). It has likely been involved in several cases of adaptive radiation, where inter-specific reproductive barriers are not completely effective (see Salzburger *et al.* 2002; Seehausen 2004; Herder *et al.* 2006). In such systems, hybrids may exhibit novel and/or extreme phenotypes trough transgressive segregation relative to parental lines, allowing the exploitation of new ecological niches and ultimately leading to speciation (Rieseberg *et al.* 1999; Schliewen & Klee 2004; Seehausen 2004; Bell & Travis 2005; Mallet 2007).

The identification and characterization of hybrids has traditionally been based on a morphological approach (Neff & Smith 1979; Allendorf *et al.* 2001). However, the use of morphology alone as a hybrid diagnostic has proven problematic because (i) samples of known hybrids are rarely available (Neff & Smith 1979) and (ii) hybrid status has been usually given on the basis of intermediacy among the morphological characters distinguishing

the parental species (e.g. Hubbs 1955; Hayden *et al.* 2010), a prerequisite that is not necessarily followed both in natural and experimental conditions (Neff & Smith 1979; Ross & Cavender 1981; Ferguson & Danzmann 1987). This is all the more the case since morphological intermediates are not necessarily hybrids (Campton 1987; Dowling & Secor 1997; Allendorf *et al.* 2001). The development of molecular genetic techniques and Bayesian clustering approaches (Pritchard *et al.* 2000; Excoffier & Heckel 2006) has greatly improved the characterization of hybrids and has shed new light on evolutionary processes involved in hybridization in the wild (Scribner *et al.* 2001; Schwenk *et al.* 2008; Hayden *et al.* 2010).

Incidence of hybridization might blur species boundaries, thus contributing to taxonomic confusion and further complicating the establishment of reliable and diagnostic, molecular and morphological characters to identify taxa. Such issues are of conservation concern, since the efficient detection and characterization of wild and admixed populations is a central concern in (i) formulating adequate management programs that are focused on the identification and protection of native gene pools and (ii) preventing hybridization between natural and stocked populations (Allendorf *et al.* 2001; Sanz *et al.* 2009).

Pupfishes of the genus *Orestias* (Teleostei: Cyprinodontidae) are endemic to the inter-Andean basin in South America (Parenti 1984b; Vila *et al.* 2007) and represent a unique case of freshwater teleostean diversification at high altitude. Given its great ecological and morphological diversity, the genus *Orestias* has been considered a 'species flock' (Villwock 1962; Kosswig & Villwock 1964; but see Parenti 1984a), where taxonomic delimitations among closely related species remain hotly debated (Villwock 1986; Loubens 1989; Lauzanne 1992; Müller 1993; Villwock & Sienknecht 1995; Villwock & Sienknecht 1996; Lüssen *et al.* 2003; Esquer Garrigos *et al*. subm.).

Knowledge on the factors and mechanisms underlying the morphological diversification of the genus is still fragmentary. Some authors have suggested that *Orestias* may have gone trough an important diversification phase during the Pleistocene (Parenti 1981; Lauzanne 1982; Parenti 1984b; Villwock 1986; Lüssen *et al.* 2003; Maldonado *et al.* 2009), likely processing as an event of adaptive radiation (Lauzanne 1982; Villwock 1986; Lüssen *et al.* 2003; Maldonado *et al.* 2009). In such a context, it is conceivable that hybridization has played a role in promoting the morphological diversification of the genus and, as a consequence, in maintaining long-standing difficulties to delimit *Orestias* species.

In this study, we focus on two species of Andean pupfishes, *O. agassizii* and *O. luteus,* which belong to two different species complexes (Lüssen *et al.* 2003; Esquer Garrigos *et al*. subm.) known to occur in sympatry in Lake Titicaca and other lagoons within the Titicaca basin (e.g. Umayo and Arapa lagoons; Parenti 1984b), but also, as recently discovered, in Lake Uru Uru (Poopó basin; C. Ibañez personal communication). *Orestias agassizii* is the most widespread, abundant and polytypic species within the genus (Parenti 1984b). It can be found in both littoral and pelagic zones of lacustrine environment, and has a generalist diet, feeding on algae, crustaceans and cladocerans (littoral), and zooplanckton (pelagic) (Lauzanne 1982; Maldonado *et al.* 2009). The species is characterized by a relatively elongated body and a very irregular head squamation pattern (Lauzanne 1982; Parenti 1984b). On the other hand, *O. luteus* is mostly restricted to Lake Titicaca (Lüssen *et al*. 2003; C. Ibañez personal communication). The species is found in the benthic zone and feeds almost exclusively on mollusks and arthropods (Lauzanne 1982; Maldonado *et al.* 2009). It is a large and robust *Orestias,* characterized by a long and wide head (Lauzanne 1982; Parenti 1984b; Maldonado *et al.* 2009).

The two species are of significant interest for assessing the impact of hybridization on an evolutionary and a conservation point of view, since (i) they are clearly separated in terms of morphological, ecological and genetic 'niches' (Lauzanne 1982; Lüssen *et al*. 2003; Maldonado *et al*. 2009; Esquer Garrigos *et al*. subm.), (ii) hybridization has been reported in both natural (Tchernavin 1944) and experimental conditions (Villwock & Sienknecht 1995; Azpiazu 2002), (iii) both species are considered "vulnerable" in Bolivia (Van Damme *et al*. 2009) and (iv) both species constitute an important component of the native fisheries (Treviño *et al.* 1992; Capriles *et al.* 2008) and are subject to breeding programs focused on restocking Lake Titicaca (Treviño *et al.* 1992).

Here we use an integrative approach combining mitochondrial (control region and cytochrome *b*) and nuclear (rhodopsin) nucleotide sequences, microsatellites, morphological characters and morphometrics to assess the incidence of natural hybridization between *O. agassizii* and *O. luteus*. Our objective is to characterize the potential hybrids originating from natural crossing between the two species, in order to provide tools to fisheries managers and actors to identify hybrid pools. In a second step, we discuss the possible impact of hybridization on the adaptive diversification and conservation aspects of *Orestias*.

Material and Methods

Sample collection and DNA extraction

Collection sites included Lago Pequeño (Lake Titicaca, Huatajata; 16°12'S–68°41'W), Lake UruUru (18°07'S–67°05'W) and Lake Poopó (18°07'S–67°04'W). These three localities are included in the large Andean hydrological system named TDPS (Titicaca - Desaguadero -

Poopó - Salar de Coipasa). The Lago Pequeño (Lake Titicaca basin) and the southern lakes UruUru and Poopó (Poopó basin), distant from ~255 km, are at least seasonally connected via the Desaguadero river (Desaguadero basin).

We used the species descriptions from Parenti (1984b) and Lauzanne (1982), and the teleostean collections at the Limnology Department of the Universidad Mayor de San Andres (UMSA), La Paz, Bolivia, attempting to classify among *O. agassizii* and *O. luteus*. After genetic analysis (see below), the 175 specimens collected represented 108 *O. agassizii*, 50 *O. luteus* and a series of 17 admixed individuals also showing admixed external characteristics that could not be attributed to neither species (Table S1, Supporting information). Collected specimens were preserved in 99% ethanol solution and deposited at UMSA. A small tissue sample was taken from the dorsal muscle of each specimen for genetic analysis.

Genomic DNA was extracted using the ABI Prism TM 6100 Nucleic Acid PrepStation (Applied Biosystems, Foster City, USA) following the manufacturer's recommendations.

PCR amplification and sequencing of mitochondrial and nuclear genes

PCR amplification of mitochondrial genes included a 398 bp fragment of control region (CR), using the primer pair L-Smel – H-Smel (Falk *et al.* 2003) and a 1102 bp fragment of cytochrome *b* (cyt*b*) using the primer pair GluF – ThrR (Machordom & Doadrio 2001). We also amplified a 689 bp fragment of a nuclear gene, representing the transmembrane domains I-VI of rhodopsin (RH), using the specific internal primer pair OrRhF – OrRhR (Esquer Garrigos *et al.* subm.). PCR amplifications were performed in a 25 µl final volume with ~100 ng of template DNA. PCR conditions for CR and RH follow Esquer Garrigos *et al.* (subm.). For cytochrome *b* (cyt*b*) the PCR mix included 0.1 mg/ml BSA, 1X PCR direct loading buffer

with $MgCl_2$, 0.25 mM of each dNTP, 0.2 μM each of forward and reverse primers, and 2 U *Taq* DNA Polymerase (Q-BIOgene, Illkirch, France). Cycling conditions were as follows: initial denaturation step at 94°C for 3 min, followed by 35 denaturation – annealing – extension cycles at 94°C for 45 s, 48°C for 1 min and 72°C for 1.15 min, and final extension step at 72°C for 15 min.

PCR products were sequenced at GENOSCOPE-Consortium National de Recherche en Génomique, Evry, France and EUROFINS MWG Operon, Ebersberg, Germany services using ABI 3730xl DNA Analyzer 96-capillary sequencers (Applied Biosystems). Nucleotide sequences were edited and aligned by eye using BioEdit version 7.0.9.0 (Hall 1999). The program PHASE version 2.1 (Stephens *et al*. 2001; Stephens and Donelly 2003) as implemented in DnaSP version 5 (Librado & Rozas 2009) was used to analyze the polymorphism present in RH and to attribute alleles in heterozygous individuals using default parameters. All probabilities for phase assignments were > 0.95. The two species did not share RH alleles in common and differed by seven 'diagnostic' sites. Intra-specific heterozygosity was observed in *O. agassizii*, but did not affect the inter-species diagnostic sites. No heterozygosity was detected within pure parental specimens of *O. luteus*. In order to confirm the inferred allelic phases, we cloned a sub-sample of individuals using the TOPO TA Cloning Kit for Sequencing (Invitrogen, Villebon-sur-Yvette, France). Sixteen clones were picked for each sample, and were amplified and sequenced using universal M13 primers (Invitrogen).

The obtained sequences were deposited in GenBank (Accession numbers: KC431254 - KC431778).

Microsatellite amplification and scoring

Ten polymorphic microsatellite markers were amplified using PCR conditions as described in Esquer Garrigos *et al*. (2011). They were scored on an ABI PRISM 3100 Genetic Analyzer using internal size standard GeneScan 500 LIZ (Applied Biosystems). Allele size was determined in GENEMAPPER version 3.0 (Applied Biosystems).

Genetic identification of pure parental specimens and hybrids

We used two Bayesian clustering approaches based on microsatellite data to probabilistically assign individual multilocus genotypes to pure parental and hybrid categories. The first method is implemented in STRUCTURE version 2.3.3 (Pritchard *et al.* 2000) and estimates the level of admixture of potential hybrids by calculating the proportion of the genome that is derived from each of the parental species. Clustering analyses were run using the admixture ancestry model, correlated allele frequencies and no prior for species information. Simulations were run five times independently, using a burn-in period of 100,000 sweeps and 500,000 MCMC iterations. Two genetic clusters ($K = 2$; Q_1 = *O. agassizii* and Q_2 = *O. luteus*) contributing to the gene pool of our sample set were established (data not shown) using the *post hoc* method proposed by Evanno *et al*. (2005). Several independent runs of different lengths were done for $K = 2$ in order to check for consistency among runs. Eventually, we assessed the assignment probability *q*i of each individual genotype originating from the genetic (species) clusters *Q1* and *Q2*.

The second method that we used is implemented in NEWHYBRIDS version 1.1b (Anderson & Thompson 2002), and identifies hybrid individuals from the posterior

probability (*Q*) of belonging to different pure parental or hybrid categories (across the last 2-3 generations). We assessed the probability for each individual to belong to the following discrete genotype frequency classes: pure *O. agassizii*, pure *O. luteus*, F1 hybrids, F2 hybrids, backcrosses with *O. agassizii* (BxOa), and backcrosses with *O. luteus* (BxOl). Simulations were run with a burn-in period of 100,000 sweeps and 500,000 MCMC iterations in five independent runs. Analysis was run without any prior information using uniform priors for mixing proportions and allele frequencies. We considered a posterior probability > 50% for assigning each specimen to a specific class (Vähä & Primmer 2006), or the highest posterior probability (if < 0.50).

We used HYBRIDLAB version 1.0 (Nielsen *et al.* 2006) to assess the power of the two Bayesian methods in accurately identifying admixture and classifying hybrids. The program simulates multilocus genotypes by randomly sampling alleles from two populations assuming random mating, linkage equilibrium, and neutrality. We simulated a data set of pure parental and hybrid classes from 50 reference genotypes of both *O. agassizii* and *O. luteus* selected using *Tq* threshold > 0.90 (Pritchard *et al.* 2000; Vähä & Primmer 2006). A total of 100 simulated genotypes were generated for the six parental and hybrid classes (pure *O. agassizii*, pure *O. luteus*, F1, F2, BxOa and BxOl). Then, simulated genotypes were used under the same model and run under the same conditions as previously described, in order to estimate threshold value for individual assignment and hybrid detection (STRUCTURE) and to assess efficiency in allocating individuals to their *a priori* known classes (NEWHYBRIDS).

The pattern of genetic differentiation between the parental individuals and identified hybrids was explored using factorial correspondence analysis (FCA) in GENETIX version 4.05 (Belkhir *et al.* 2004).

Genetic diversity estimates and diagnostic genetic tools

Estimates of levels of diversity using nucleotide sequences (mt- and nDNA) and microsatellite data were restricted to the two pure parental groups, *O. agassizii* and *O. luteus*. For nucleotide sequences, we used DnaSP version 5 to calculate the number of haplotypes (*h*), haplotype diversity (*Hd*), number of polymorphic sites (*S*), nucleotide diversity per site (*Pi*), and overall mean distances between haplotypes.

Incidence of null alleles in microsatellites and scoring errors due to stuttering or large allele drop-out was assessed using MICRO-CHECKER version 2.2.3 (Oosterhout *et al.* 2004). Departures from Hardy-Weinberg proportions, and linkage disequilibrium were estimated in Genepop version 1.2 (Raymond & Rousset 1995). To minimize type-I errors, *P* values were adjusted for multiple test of significance using the sequential Bonferroni correction at the 5% nominal level (Rice 1989). Genetic diversity at each microsatellite locus was quantified in GenAlEx version 6.0 (Peakall & Smouse 2006) as the number of alleles per locus (*NA*), number of private alleles (*NPA*), observed (*H*O) and expected heterozygosity (*H*E). To account for differences in diversity due to differences in sample sizes, allelic richness and private allele richness were recalculated under the rarefaction method implemented in HPRARE version 1.0 (Kalinowski 2005).

Morphological analyses

Morphological discrimination between *O. agassizii*, *O. luteus* and potential hybrids was assessed using 17 standard morphometric variables (see Table 5) measured by digital or dial calipers (nearest 0.1 mm).

Groups were designed *a priori* from the Bayesian genetic clustering results (see above). Using a MANOVA test in SPSS version 16.0 (Levesque 2007), we detected a significant sexual dimorphism in *O. agassizii* (Wilks' $\lambda = 0.488$; $P = 0.000$). Thus, multivariate analyses were conducted separately for each sex in both species.

We eliminated the influence of size on shape variation before comparing groups using the method of "log-shape ratio" (Mosimann 1970; Mosimann & James 1979). Raw data were log transformed and for each specimen an arithmetic mean was calculated for all the variables. The arithmetic mean was then subtracted to each of the log transformed variables. The adjusted values (log-shape ratios) were subject to a discriminant analysis (DA) i) to investigate the pattern of differentiation between both species and their putative hybrids, ii) to identify the most discriminative morphometric characters, and iii) to assess the percentage of specimens correctly classified into their respective groups using the jackknife method.

Assessment of Lauzanne's (1982) identification criteria for O. agassizii and O. luteus

On the basis of our pure parental and hybrid groups genetic assignment (see above), we re-assessed the consistency of a short series of ratios (morphometric characters) and one meristic trait reported by Lauzanne (1982) to discriminate between *O. agassizii* and *O. luteus*. Ratios included (i) head length / standard length (B/A), (ii) head width / standard length (N/A), and (iii) body height / standard length (F/A). The meristic character concerns the type of scales on the head, which are thickened and granular in *O. luteus* and non-granular in *O. agassizii*.

Results

Identification of purebred individuals and hybrids

Simulated data

Results from HYBRIDLAB showed that 100% of simulated pure parental genotypes could be correctly assigned in STRUCTURE using T*q* = 0.86 (Fig. 1). All the simulated F1 hybrids had assignment probabilities below the T*q* threshold (0.354–0.646). The picture was similar for 98% of the simulated F2 hybrids (0.067 – 0.093). Simulated backcrosses were assigned to pure parental classes in 13% (BxOa) and 12% (BxOl) of the cases. Assignments in NEWHYBRIDS using HYBRIDLAB simulations required *Tq* = 0.66 to reach 100% of correct allocation to pure parental *O. agassizii*, *O. luteus* and F1 hybrids. A great proportion of simulated individuals were misclassified within F2 hybrids (74%), and backcrosses BxOa (27%) and BxOl (32%).

Observed data

In STRUCTURE, pure parental *O. agassizii* and *O. luteus* assignment probabilities averaged *Q*1 = 0.993 (range: 0.953–0.997) and *Q*2 = 0.992 (range: 0.972–0.996) (Table 1). NEWHYBRIDS consistently attributed to their own class the pure parental specimens of *O. agassizii* and *O. luteus* as assigned by STRUCTURE (Table 1). Admixed specimens were identified at the two sampling sites where both parental species were collected (i.e. Huatajata and lake Uru Uru; see Fig. 2). With STRUCTURE, 16 out of the 17 morphologically admixed individuals were identified as genetically admixed (Fig. 1, Table 1; *qi* range: 0.253–0.747). The last morphologically admixed individual (2U394) showed an assignment probability to *O. luteus* (*qi* = 0. 88) slightly above the threshold established from simulations. The same

series of 16 specimens identified as admixed by STRUCTURE was classified as F1 hybrids by NEWHYBRIDS since they showed inferred ancestry probabilities in majority attributed to the F1 hybrid class (Table 1). The specimen 2U394 had a similar probability of belonging to a pure parental *O. luteus* or a backcross with this species (BxOl); additional evidence based on RH heterozygosity classifies the individual as a backcross (see below).

The factorial correspondence analysis (FCA) showed a clear distinction among pure parental *O. agassizii* and *O. luteus,* and the hybrids identified using microsatellite-based assignment procedures (Fig. 3). Specimen 2U394 placed near to the *O. luteus* group.

Genetic diversity estimates and diagnostic genetic tools

Genetic diversity indices estimated from nucleotide sequences were generally higher for *O. luteus* than for *O. agassizii*, although *O. agassizii* showed a higher diversity for cytochrome *b* (Table 2). Both cyt*b* and CR showed sites diagnostic for *O. agassizii* and *O. luteus* (67 and 26, respectively; data not shown). Among the 17 putative hybrids (including specimen 2U394), 13 had *O. luteus* mtDNA (Table S1, Supporting information). The seven diagnostic sites in RH nucleotide sequences discriminating between *O. agassizii* and *O. luteus* (Table 3) were heterozygous in the 16 specimens identified as hybrids and in the BxOl backcross 2U394 using microsatellite-based assignment methods (Table S1, Supporting information).

The expected distribution of homozygote size classes calculated with MICRO-CHECKER suggested the incidence of null alleles for loci A106, A116, C101 and D110 in *O. agassizii*, and loci A116 and C105 in *O. luteus*. Given that no significant Hardy-Weinberg departure was detected for these loci in *O. agassizii* after sequential Bonferroni correction ($P > 0.006$), except for locus A116 in *O. luteus* ($P < 0.006$), we considered all loci in our

Bayesian inferences (Table 4). Locus A116 was also retained because it showed a rather non-overlapping amplification size between both species (see below).

Exact test of linkage disequilibrium revealed significant association between loci B1 and A116 in *O. agassizii* (P = 0.000). No linkage disequilibrium between pairs of loci was detected for *O. luteus*.

Microsatellite diversity was generally higher for *O. agassizii* than *O. luteus*. The number of alleles in *O. agassizii* and *O. luteus,* respectively, averaged 19.3 (range: 5–34) and 12.2 (range: 2–30); the number of private alleles averaged 11.5 (range: 2.8–25.75) and 4.4 (range: 1–12); mean observed and expected heterozygosities averaged 0.698 (range: 0.141–0.88) - 0.746 (range: 0.168–0.954) and 0.653 (range: 0.12–0.94) - 0.698 (0.11–0.96). Alleles found in loci C102, A9a and B104 were private for each species (with no allele size overlap in C102). Locus A116 had only one allele shared between the two species (Table 4, Fig. 4).

Morphological analyses

Standard length range was (i) 88.74–118.27 mm (mean = 101.47) for males and 78.45–150.2 mm (mean = 102.62) for females in *O. agassizii*, (ii) 81.39–118.8 mm (mean = 98.58) for males and 76.39–116.0 mm (mean = 98.21) for females in *O. luteus*, and (iii) 100.53–138.77 mm (mean = 116.36) for males and 47.92–129.93 mm (mean = 107.18) for females in hybrids.

For both sexes, the discriminant analysis showed that the two parental species (*O. agassizii* and *O. luteus*) and their hybrids formed three distinct groups. Hybrids were closer to the parental species *O. agassizii* on the first axis (Fig. 5). The first and second discriminant functions (CF1 and CF2) were significant ($p < 0.05$) and contributed to 91.3% and 8.7% (males) and 95.3% and 4.7% (females) of the variance among groups, respectively. CF1 was

mainly associated with head width in both sexes. In males, CF2 was mainly associated with body width (Table 5).

Cross-validated classification showed that 93% of females and 85% of males were correctly classified (Table 6). Within the *O. agassizii* group, six female and five male specimens were misclassified in the hybrid group, and one male in the *O. luteus* group. Within the hybrid group, one female and three male specimens were misclassified as *O. agassizii*, and one female as *O. luteus*. No specimen from the *O. luteus* group was misclassified.

Lauzanne's (1982) identification criteria for O. agassizii and O. luteus

The values of three morphometric ratios reported by Lauzanne (1982) to discriminate between *O. agassizii* and *O. luteus* overlapped between the two pure parental groups, and also between hybrids and parental groups (Table 7). Within hybrids, standard deviation was greater than in pure parental groups. The meristic character related to the type of scales on the head clearly discriminated between pure parental groups. Hybrids showed in their majority (16 out of 17) thickened and non-granulated scales, characteristic of *O. agassizii* (see Table S1, Supporting information).

Discussion

Integrative diagnosis of Orestias agassizii x O. luteus hybrids

Mt- and nDNA sequences, microsatellites and morphological characters (meristic and morphometric) were useful in differentiating between the parental species *O. agassizii and O. luteus*, whereas the morphological diagnostic ratio values provided by Lauzanne (1982) proved overlapping. For the first time on a genetic basis, our integrative approach confirmed the incidence of natural hybridization between two species of *Orestias*. We showed that heterozygosity at species diagnostic sites of RH sequences constituted an effective diagnostic tool for detecting *O. agassizii x O. luteus* hybrids (RH carrying the copies of the two parental species), further confirming the usefulness of RH in hybrid detection within vertebrates (Vogel & Johnson 2008; Collins 2012). Since our RH-based identification of hybrids was largely confirmed by assignment procedures using recently developed microsatellites for *O. agassizii* and *O. luteus* (Esquer Garrigos *et al.* 2011), we consider that we here provide a genetic diagnostic 'kit' (RH + 10 microsatellites) readily usable to detect hybrids between the two species.

Clustering analyses based on 10 microsatellite loci proved efficient in accurately identifying pure parental and first-generation hybrid (F1) categories, but resulted in more problematic assignment of second-generation hybrids (F2) and backcrosses (Fig. 1), as expected given the relatively low number of loci considered for the study (see Anderson & Thompson 2002; Vähä & Primmer 2006). A better characterization of hybridization events in *Orestias* will require the development of additional loci, all the more since preliminary assessment of microsatellite genetic distances among the four species complexes of *Orestias*

(*agassizii, luteus, gilsoni, cuvieri*; Esquer Garrigos *et al*. subm.) shows low levels of divergence (Y. Esquer Garrigos and P. Gaubert, personal observations). Within our sample set, one backcross with *O. luteus* was detected (specimen 2U394; Table 1 and Appendix 1). Unexpectedly, this specimen was assigned by the discriminant analysis to the *O. agassizii* morphological group (Fig. 5 and Table S1, Supporting information). Even if our results are preliminary, such a discrepancy between genotype and phenotype assignment could reflect the incidence of several backcrossed generations involving the two species (see Burgarella *et al.* 2009), further complicating assignment procedures within *Orestias*, at least in the geographic area considered.

Deviation from intermediate morphotypes in F1 hybrids has been detected among various freshwater teleosteans in both natural and laboratory conditions (Ross & Cavender 1981; Campton 1987; Ferguson & Danzmann 1987). In our case, the *O. agassizii x O. luteus* hybrids deviated from morphological intermediacy between the parental species, and were closer to *O. agassizii*, mostly in terms of head width (Fig. 5). Although the discriminant analysis distinguished hybrids from their parental groups, several hybrids were misclassified, mostly as *O. agassizii* (Table 6, Table S1, Supporting information). Following this same picture, a majority of hybrids exhibited the meristic character trait characterizing *O. agassizii* (head with thickened and non-granulated scales; Table S1, Supporting information). Thus, our results suggest that (i) the intermediacy criterion might be inapplicable to detect *Orestias* hybrids on a morphological basis, and (ii) disruptive genetic and regulatory interactions may impact on the morphogenesis of F1 hybrids (see Ferguson and Danzmann 1987; Ross and Cavender 1981). Finally, the morphometric ratios proposed by Lauzanne's (1982) to discriminate between *O. agassizii* and *O. luteus* within Lago Pequeño seemed to have overlapping ranges using our genetically delimited species groups (Table 7). Thus, we

showed that the two species had a greater morphological variability than expected, when sampling into larger series of specimens and geographic ranges (although we cannot exclude the impact of allometry on our results given the heterogeneous series of maturity stages considered; Table S1, Supporting information). We also evidenced that hybrids showed a greater range in the morphometric ratios than the parental groups, rejoining the view that hybridization generates a wide spectrum of morphological variants difficult to categorize, even as hybrids (Neff and Smith 1979). Beyond the hybrid issue, our results illustrate the difficulties in morphologically delimiting among *Orestias* species (Villwock 1986; Loubens 1989; Lauzanne 1992; Müller 1993; Villwock and Sienknecht 1995; Villwock and Sienknecht 1996; Lüssen *et al*. 2003; Esquer Garrigos *et al*. subm.), and thus highlight the need to better characterize morphological variability within the genus.

Incidence of and potential mechanisms related to hybridization between O. agassizii and O. luteus

The rate of natural hybridization seems more frequent in teleosteans than in any other group of vertebrates (Schwenk *et al*. 2008). Average hybridization levels in animals have been estimated between c. 1 (Mallet 2005) and 10% (Schwenk *et al*. 2008). Rates of hybridization in freshwater teleosteans under natural conditions have been reported to range from 30% in the Cyprinidae, to 8% in the Salmonidae, only 3% in the Cichlidae and 2% in Cyprinodontidae (Scribner *et al*. 2001). The 10% rate of hybridization that we report between *O. agassizii* and *O. luteus* is thus c. five-fold greater than the value reported for Cyprinodontidae (Scribner *et al*. 2001), and can be considered high for a vertebrate taxon. Since hybrids were identified both in Lake Titicaca and Lake Uru Uru, hybridization may be

widespread and take place wherever both species are sympatric. Our results are supported by captive-breeding data that showed the successful hybridization between *O. agassizii* and *O. luteus* collected from one of our study sites (Huatajata, Lake Titicaca), reaching viable F2 hybrids (Villwock & Sienknecht 1995). Altogether, the combination of empirical and experimental evidences suggests weak reproductive isolation barriers between those two morphologically and ecologically differentiated species of *Orestias* (see Lauzanne 1982; Maldonado *et al*. 2009). Indeed, hybrid specimens have also been reported between *O. agassizii* and *O. pentlandii* (Autoridad binacional autónoma del sistema hídrico T.D.P.S 2002) and *O. cuvieri* and *O. pentlandii* (Tchernavin 1944). Such observations within the genus *Orestias* fit the evolutionary framework of a recent or ongoing adaptive radiation, where reproductive isolation mechanisms among species are frequently weak or absent, and the viability and fertility of F1 hybrids is high (Wood & Foote 1990; Hatfield & Schluter 1999).

One may wonder how hybridization takes place between two species exploiting supposedly different ecological niches, such as *O. agassizii* and *O. luteus*. The observation of mtDNA distribution showed that 13 hybrids (77%) had a haplotype belonging to *O. luteus*. If we hypothesize that most of them were F1 hybrids and that selective pressure is similar for each type of cross (but see below), a first hypothesis is that unbalanced relative abundances in favor of *O. agassizii* (especially males) promotes asymmetric hybridization between *agassizii* males and *luteus* females (e.g. McGowan & Davidson 1992; Lepais *et al.* 2009). Unfortunately, abundance estimates for the two species are not available. Another hypothesis that would potentially explain hybridization between *O. agassizii* and *O. luteus* is the combination of weak pre-zygotic reproductive barriers with similar breeding temporality (throughout the year; Lauzanne 1992) and similar spawning behavior —i.e., lying eggs in

shallow waters on the submerged vegetation belt (Bustamante & Treviño 1980). We posit a third hypothesis relying on laboratory observation evidence, where individuals of both species showing a juvenile external appearance were sexually mature (Villwock & Sienknecht 1995). Since juveniles of *O. agassizii* and *O. luteus* are reported shoaling together in the same habitats (Villwock & Sienknecht 1996; C. Ibañez, personal observations), it is possible that sexually mature, but phenotypically and ecologically juvenile-like populations from both species hybridize.

Detecting hybridization is relevant for apprehending evolutionary processes within the genus *Orestias*. Several studies have shown that hybrids could have an equal or a greater fitness ('hybrid vigor') than parental species (Arnold & Hodges 1995; Barton 2001; Rosenfield *et al*. 2004). Evidence for such interspecific hybrid vigor has been provided in a wide variety of fisheries breeding programs targeted at enhancing traits such as growth rate and/or parasite resistance (Bartley *et al.* 2001). In *Orestias*, Aspiazu (2000) reported that F1 hybrids between males *O. agassizii* and females *O. luteus* showed the highest fitness in terms of growth and survival rates, even compared to the two parental species. Interestingly, the majority of the hybrids that we detected are likely to belong to this cross category. Although our results are preliminary, they suggest that *O. agassizii* ♂ x *O. luteus* ♀ hybrids could have a higher fitness —expressed as survival rate, for instance— in natural conditions too. We also showed that F1 hybrids (including the potential backcross with *O. luteus*) were morphologically closer to the parental species *O. agassizii*. Such a trend may reflect the incidence of natural selection in favor of the *O. agassizii* morphotype, which is the most ecologically successful (i.e., widespread and diversified) within the genus (Lauzanne 1982; Parenti 1984b), but we cannot rule out the role of developmental constraints. Eventually, hybridization may boost diversification in the context of adaptive radiation by enhancing

genetic variability (Rieseberg *et al.* 1999; Seehausen 2004; Mallet 2005; Mallet 2007), increasing divergence among species in the initial stage of the radiation (Schliewen & Klee 2004; Bell & Travis 2005), and promoting evolutionary innovations ('key innovations'; Seehausen 2004) through transgressive segregation (Riesenberg *et al*. 1999). Further studies on hybridization and hybrid fitness in *Orestias* will be crucial to understand the past and likely ongoing diversification processes within the genus, notably focusing on how reproductive isolation and morphotypes are maintained or consolidated in the face of interspecific gene flow (see Arnold & Hodges 1995; Seehausen 2004; Mallet 2005).

Implications for the conservation and management of Orestias species

Because they represent an important component of the native Andean fisheries (Treviño *et al.* 1992; Capriles *et al.* 2008) and are regionally classified as vulnerable (Van Damme *et al*. 2009), *O. agassizii* and *O. luteus* are species of economic and conservation concern. Moreover, the widespread *O. agassizii* may occupy very restricted and fragile habitats that are sensitive to eutrophication and pollution (Northcote 1992), such as isolated lagoons, seasonal wetlands and small freshwater ponds within salt pans (‘salares’). Hydro-acoustic evaluations have shown steady stock reduction (Vila *et al*. 2007) due to several human-induced factors, combining overfishing (Orlove *et al.* 1992), introduction of invasive species such as the rainbow trout (*Oncorhynchus mykiss*) and the silverside (*Odontesthes bonariensis*) (Lauzanne 1992, Vila *et al*. 2007), the replacement of traditional fishing techniques by highly extractive methods (Vellard 1992), environmental degradation trough water disposal (notably for mining) (see Northcote 1992), and over-extraction of the totora reed (*Schoenoplectus tatora*), an aquatic plant constituting the habitat of several *Orestias* species.

Facing this situation, several organisms from Peru and Bolivia have carried out research programs on artificial spawning with the aim of reinforcing and restocking local populations in Lake Titicaca through the release of captive-bred specimens (Treviño *et al.* 1992, Aspiazu 2000). The diagnosis of pure parental and/or hybrid groups is a key point in such a conservation issue relying on the captive-breeding of wild specimens (Rhymer & Simberloff 1996; Allendorf *et al.* 2001; Sanz *et al.* 2009). Our study provides evidence that hybridization occurred at a noticeable rate (~10%) and wherever *O. agassizii* and *O. luteus* were found in sympatry (independent of the type of habitats), thus suggesting that hybridization between the two species may be a widespread phenomenon across Lake Titicaca. In order to avoid the release of genetically unfit specimens and the risk of outbreeding depression (Miller *et al.* 2004; Grewe *et al.* 2007), we recommend that (i) all captive-bred populations are screened for *O. agassizii x O. luteus* hybrids using our genetic diagnostic 'kit' (RH + 10 microsatellites) before release, and (ii) the same process is applied to wild populations before sampling for captive breeding.

The detection of hybridization in 'undisturbed' environmental conditions between *O. agassizii* and *O. luteus* is also relevant for future scenarios of conservation and evolution in Andean killifishes. Indeed, hybridization in freshwaters may occur after human-induced ecological perturbations, leading to eutrophication and high water turbidity (e.g. Witte *et al.* in press). Given the anthropic pressure that the habitats of *Orestias* are subjected to (see above) and the likely disturbing effect of future climate changes on Andean freshwater resources (Kundzewicz *et al.* 2008), there is a genuine opportunity for hybridization events to increase in numbers and sites among Andean killifishes (see Scribner *et al.* 2001). On an evolutionary point of view, hybridization due to human-induced environmental stress may constitute a source of incipient diversification in freshwater ecosystems (Mallet 2007; Witte *et*

al. in press), although not likely to compensate the rate of species extinction due to such disturbances (Vonlanthen *et al.* 2012). Concerning the two sites where we detected natural hybridization between *O. agassizii* and *O. luteus* (Lago Pequeño and Lake Uru Uru), we recommend that they are considered as potential evolutionary hotspots (hybrid swarm theory; Seehausen 2004) that should deserve a specific conservation attention.

Acknowledgments

We thank Alexander Flores, Kelvin Herbas and Ramón Catari for field assistance (UMSA), to Claude Ferrara (Direction des Collections, MNHN) for taking photographs of *Orestias*, and Thierry Oberdorff (UMR BOREA, MNHN) for early stage project management. Our study was founded by the Agence Nationale de la Recherche (project "FISHLOSS" – ANR-09-PEXT-008). Laboratory work was done at the 'Service de Systématique Moléculaire' of the Muséum National d'Histoire Naturelle (USM CNRS 2700), and was supported by the 'Consortium National de Recherche en Génomique', and the 'Service de Systématique Moléculaire' of the Muséum National d'Histoire Naturelle (CNRS-UMS 2700). It is part of agreement 2005/67 between Genoscope and the Muséum National d'Histoire Naturelle on the project "Macrophylogeny of life" directed by Guillaume Lecointre, and by the project @SPEED-ID "Accurate SPEcies Delimitation and Identification of eukaryotic biodiversity using DNA markers" proposed by F-Bol, the French Barcode of life initiative. YEG received a PhD-fellowship of the Institut de Recherche pour le Développement-Département Soutien et Formation (IRD-DSF).

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N *et al*. (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution*, **16**, 613–622.

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160**, 1217–1229.

Arnold ML (1992) Natural hybridization as an evolutionary process. *Annual Review of Ecology and Systematics*, **23**, 237–261.

Arnold ML, Hodges SA (1995) Are natural hybrids fit or unfit relative to their parents? *Trends in Ecology and Evolution*, **10**, 67–71.

Autoridad binacional autónoma del sistema hídrico T.D.P.S (2002) Subcontrato no. 21.24: Desarrollar la capacidad de programas de pesca artesanal en el ambito peruano del sistema T.D.P.S (ed IIP Qollasuyo-CIPP Chucuito), 263 p. UNA Puno, Peru.

Azpiazu WI (2002) Hibridación artificial entre carachi *(Orestias agassii)* y punku *(Orestias luteus)* "killifish" del Lago Titicaca, Tesis de Licenciatura, Facultad de Agronomía, Universidad Mayor de San Andrés, La Paz, Bolivia.

Bartley DM, K R, Immink AJ (2001) The use of inter specific hybrids in aquaculture and fisheries. *Reviews in Fish Biology and Fisheries*, **10**, 325–337.

Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology*, **10**, 551–568.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, Logiciel Sous Windows TM Pour la Génétique des Populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

Bell MA, Travis MP (2005) Hybridization, transgressive segregation, genetic covariation, and adaptive radiation. *Trends in Ecology and Evolution*, **20**, 358–361.

Burgarella C, Lorenzo Z, Jabbour-Zahab R, Lumaret R, Guichoux E, Petit RJ, Soto A, Gil L (2009) Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q. ilex*). *Heredity*, **102**, 442–452.

Bustamante E, Treviño H (1980) Descripción de las pesquerías en el Lago Titicaca 1975–1979, 73 p. Instituto Mar, Puno, Perú.

Campton DE (1987) Natural hybridization and introgression in fishes: methods of detection and genetic interpretations. In: *Population genetics and fishery managemen*t (eds Ryman N & Utter F), pp. 161–192. University of Washington Press, Seattle, WA.

Capriles JM, Domic AI, Moore KM (2008) Fish remains from the Formative Period (1000 BC–AD 400) of Lake Titicaca, Bolivia: Zooarchaeology and taphonomy. *Quaternary International*, **180**, 115–126.

Collins RA (2012). Investigating interspecific hybridisation in ornamental fishes. *Figshare electronic publication*. doi: 10.6084/m9.figshare.96149.

Dowling TE, Secor CL (1997) The role of hybridization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics*, **28**, 593–619.

Esquer Garrigos Y, Lambourdière J, Ibañez C, Gaubert P (2011) Characterization of ten polymorphic microsatellite loci in the Andean pupfish *Orestias agassizii*, with cross-amplification in the sympatric *O. luteus*. *Conservation Genetics Resources*, **3**, 17–19.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics*, **7**, 745–758.

Falk TM, Teugels GG, Abban EK, Villwock W, Renwrantz L (2003) Phylogeographic patterns in populations of the black-chinned tilapia complex (Teleostei, Cichlidae) from coastal areas in West Africa: support for the refuge zone theory. *Molecular Phylogenetics and Evolution*, **27**, 81–92.

Ferguson M, Danzmann R (1987) Deviation from morphological intermediacy in interstrain hybrids of rainbow trout, *Salmo gairdneri*. *Environmental Biology of Fishes*, **18**, 249–256.

Grewe PM, Patil JG, McGoldrick DJ, Rothlisberg PC, Whyard S, Hinds LA, Hardy CM, Vignarajan S, Thresher RE (2007) Preventing genetic pollution and the establishment of feral populations: a molecular solution. In: *Ecological and genetic implications of aquaculture activities. Reviews: Methods and Technologies in Fish Biology and Fisheries* (ed T.M. Bert), pp. 103–114. Springer, Dordrecht, The Netherlands.

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hatfield T, Schluter D (1999) Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution*, **53**, 866–873.

Hayden B, Pulcini D, Kelly-Quinn M, O'Grady M, Caffrey J, McGrath A, Mariani S (2010) Hybridization between two cyprinid fishes in a novel habitat: genetics, morphology and life-history traits. *BMC Evolutionary Biology*, **10**, 169.

Herder F, Nolte AW, Pfaender J, Schwarzer J, Hadiaty RK, Schliewen UK (2006) Adaptive radiation and hybridization in Wallace's Dreamponds: evidence from sailfin silversides in the Malili Lakes of Sulawesi. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 2209–2217.

Hubbs CL (1955) Hybridization between fish species in nature. *Systematic Zoology*, **4**, 1–20.

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187–189.

Kosswig C, Villwock W (1964) Das problem der intralakustrischen speziation im Titicaca- und im Lanaosee. *Verhandlungen der Deutschen Zoologischen Gesellschaft*, **28**, 95–102.

Kundzewicz ZW, Mata LJ, Arnell NW, Döll P, Jimenez B, Miller K, Oki T, Şen Z, Shiklomanov I (2008) The implications of projected climate change for freshwater resources and their management. *Hydrological Sciences Journal*, **53**, 3–10.

Lauzanne L (1982) Les *Orestias* (Pisces, Cyprinodontidae) du Petit lac Titicaca. *Revue d'Hydrobiologie Tropicale*, **15**, 39–70.

Lauzanne L (1992) Fish fauna, native species: the *Orestias*. In: *Lake Titicaca: a synthesis of limnological knowledge* (eds Dejoux C & Iltis A), pp. 405–419, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Lepais O, Guichoux E, Lavabre JE, Alberto F, Kremer A, Gerber S (2009) Species relative abundance and direction of introgression in oaks. *Molecular Ecology*, **18**, 2228–2242.

Levesque R (2007) SPSS Programming and Data Management: A Guide for SPSS and SAS Users, Fourth Edition, SPSS Inc., Chicago IL.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Loubens G (1989) Observations sur les poissons de la partie Bolivienne du lac Titicaca. IV. *Orestias* spp., *Salmo gairdneri* et problèmes d'aménagement. *Revue d'Hydrobiologie Tropicale*, **22**, 157–177.

Lüssen A, Falk T, Villwock W (2003) Phylogenetic patterns in populations of Chilean species of the genus *Orestias* (Teleostei: Cyprinodontidae): Results of mitochondrial DNA analysis. *Molecular Phylogenetics and Evolution*, **29**, 151–160.

Machordom A, Doadrio I (2001) Evidence of a cenozoic Betic-Kabilian connection based on freshwater fish phylogeography (*Luciobarbus*, Cyprinidae). *Molecular Phylogenetics and Evolution*, **18**, 252-263.

Maldonado E, Hubert N, Sagnes P, De Mérona B (2009) Morphology-diet relationships in four killifishes (Teleostei, Cyprinodontidae, *Orestias*) from Lake Titicaca. *Journal of Fish Biology*, **74**, 502–520.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, **20**, 229–237.

Mallet J (2007) Hybrid speciation. *Nature*, **446**, 279–283.

McGowan C, Davidson WS (1992) Unidirectional natural hybridization between brown trout (*Salmo trutta*) and Atlantic salmon (*S. salar*) in Newfoundland. *Canadian journal of fisheries and aquiatic sciences*, **49**, 1953–1958.

Miller LM, Close T, Kapuscinski R (2004) Lower fitness of hatchery and hybrid rainbow trout compared to naturalized populations in Lake Superior tributaries. *Molecular Ecology*, **13**, 3379–3388.

Mosimann JE (1970) Size allometry: size and shape variables with characterizations of the lognormal and generalized gamma distributions. *Journal of the American Statistical Association*, **65**, 930–945.

Mosimann JE, James FC (1979) New statistical methods for allometry with application to florida red-winged blackbirds. *Evolution*, **33**, 444–459.

Müller R (1993) Critical remarks on the revision of the genus *Orestias* (Pisces Cyprinodontidae) by Parenti (1984). *Zoologische Jahrbücher. Abteilung für Anatomie und Ontogenie der Tiere*, **123**, 31–58.

Neff NA, Smith GR (1979) Multivariate analysis of hybrid fishes. *Systematic Zoology*, **28**, 176–196.

Nielsen AEG, Bach LA, Kotlicki P (2006) HYBRIDLAB ver. 1.0: A program for generating simulated hybrids from population samples. *Molecular Ecology Notes*, **6**, 971–973.

Northcote TG (1992) Eutrophication and pollution problems. In: *Lake Titicaca: a synthesis of limnological knowledge* (eds Dejoux C, Iltis A), pp. 551–559. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Oosterhout C, Hutchinson W, Wills D, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes*, **4**, 535–538.

Orlove BS, Levieil DP, Treviño HP (1992) Social and economic aspects of the fisheries. In: *Lake Titicaca: a synthesis of limnological knowledge* (eds Dejoux C & Iltis A), pp. 500–504. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Parenti LR (1981) A phylogenetic and biogeographic analysis of cyprinodontiform fishes (Teleostei, Atherinomorpha) *Bulletin of the American Museum of Natural History*, **168**, 335–557.

Parenti LR (1984a) Biogeography of the Andean killifish genus *Orestias* with comments on the species flock concept. In: *Evolution of fish species flocks.* (eds. Echelle AA, Kornfield I), pp. 85–92. University of Maine Press, Orono, ME.

Parenti LR (1984b) A taxonomic revision of the Andean killifish genus *Orestias* (Cyprinodontiformes, Cyprinodontidae). *Bulletin of the American Museum of Natural History*, **178**, 107–214.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, **27**, 83–109.

Rice WR (1989) Analysing tables of statistical test. *Evolution*, **43**, 223–225.

Rieseberg LH, Archer MA, Wayne RK (1999) Transgressive segregation, adaptation and speciation. *Heredity*, **83**, 363–372.

Ross MR, Cavender TM (1981) Morphological analysis of four experimental intergeneric cyprinid hybrid crosses. *Copeia*, **1981**, 377–387.

Rosenfield, JA, Nolasco S, Lindauer S, Sandoval C, Kodric-Brown A (2004) The role of hybrid vigor in the replacement of pecos pupfish by its hybrids with sheepshead minnow. *Conservation Biology*, **18**, 1589–1598.

Salzburger W, Baric S, Sturmbauer C (2002) Speciation via introgressive hybridization in East African cichlids? *Molecular Ecology*, **11**, 619–625.

Sanz N, Araguas RM, Fernandez R, Vera M, Garcia-Marin JL (2009) Efficiency of markers and methods for detecting hybrids and introgression in stocked populations. *Conservation Genetics*, **10**, 225–236.

Schliewen U, Klee B (2004) Reticulate sympatric speciation in Cameroonian crater lake cichlids. *Frontiers in Zoology*, **1**, 1–12.

Schwenk K, Brede N, Streit B (2008) Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **12**, 2805–2811.

Scribner KT, Page KS, Bartron ML (2001) Hybridization in freshwater fishes: a review of case studies and cytonuclear methods of biological inference. *Reviews in Fish Biology and Fisheries*, **10**, 293–323.

Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology and Evolution*, **19**, 198–207.

Seehausen O (2006) Losing biodiversity by reversal speciation. *Current biology*, **16**, R334–R337.

Stephens M, Donelly P (2003) A comparison of Bayesian methods for haplotype reconstruction. *American Journal of Human Genetics*, **73**, 1162–1169.

Stephens M, Smith NJ, Donelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Stemshorn KC, Reed FA, Nolte AW, Tautz D (2011) Rapid formation of distinct hybrid lineages after secondary contact of two fish species (*Cottus* sp.). *Molecular Ecology*, **20**, 1475–1491.

Tchernavin VV (1944) A revision of the subfamily Orestiinae. *Proceedings of the Zoological Society of London*, **114**, 140–233.

Tobler M, Carson EW (2010) Environmental variation, hybridization, and phenotypic diversification in Cuatro Ciénegas pupfishes. *Journal of Evolutionary Biology*, **23**, 1475–1489.

Treviño H, Torres Caleron J, Roncal Gutierrez M (1992) The fishery potential. In: *Lake Titicaca. A synthesis of limnological knowledge* (eds Dejoux C & Iltis A), pp. 539–549. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Vähä J-P, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, **15**, 63–72.

Van Damme PA, Carvajal-Vallejos F, Sarmiento J, Barrera S, Osinaga K, Miranda-Chumacero G (2009) Peces. In: *Libro rojo de la fauna silvestre de vertebrados de Bolivia* (eds Ministerio de Medio Ambiente y Agua), pp. 25-90. La Paz, Bolivia.

Vellard J (1992) Former lake fisheries and fish fauna of the lake. In: *Lake Titicaca: a synthesis of limnological knowledge* (eds Dejoux C & Iltis A), pp. 495–499.Kluwer Academic Publishers, Dordrecht, The Netherlands.

Vila I, Pardo R, Scott S (2007) Freshwater fishes of the Altiplano. *Aquatic Ecosystem Health and Management Society*, **10**, 201–211.

Villwock W (1962) Die Gattung *Orestias* (Pisces, Microcyprini) und die Frage der intralakustrischen Speziation im Titicaca-Seengebiet. *Verhandlungen der Deutschen Zoologischen Gesellschaft Wien 1962, Zoologischer Anzeiger*, **26**, 610–624.

Villwock W (1986) Speciation and adaptive radiation in the Andean *Orestias* fishes. In: *High altitude tropical biogeography* (eds Vuilleumier F & Monasterio M), pp. 387–403. Oxford University Press, Oxford, UK.

Villwock W, Sienknecht U (1995) Intraspezifische variabilität im genus *Orestias* Valenciennes, 1839 (Teleostei: Cyprinodontidae) und zum problem der artidentität. *Mitteilungen aus dem Hamburgischen Zoologischen Museum und Institut*, **92**, 381–398.

Villwock W, Sienknecht U (1996) Contribution to the knowledge and history of Chilean fishes. The cyprinodontids of the genus Orestias Val. 1839 (Teleostei: Cyprinodontidae) of the Chilean Altiplano. *Medio Ambiente*, **13**, 119–126.

Vogel LS, Johnson SG (2008) Estimation of hybridization and introgression frequency in toads (genus:*Bufo*) using DNA sequence variation at mitochondrial and nuclear loci. *Journal of Herpetology*, **42**, 61–75

Vonlanthen P, Bittner D, Hudson AG, Young KA, Muller R, Lundsgaard-Hansen B, Roy D, Di Piazza S, Largiader CR, Seehausen O (2012) Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, **482**, 357-362.

Witte F, Seehausen O, Wanik JH, Kishe-Machumu A, Rensing M, Goldschmidt T (in press) Cichlid species diversity in naturally and anthropogenically turbid habitats of Lake Victoria, East Africa. *Aquatic Sciences*.

Wood CC, Foote CJ (1990) Genetic differences in the early development and growth of sympatric sockeye salmon and kokanee (*Oncorhynchus nerka*) and their hybrids. *Canadian journal of fisheries and aquatic sciences*, **47**, 2250–2260.

Author Contributions Box

Y.E.G., P.G. and B.H. conceived the study.

Y.E.G., K.K., J.L. and A.C. performed lab work.

C.Z and C.I. collected samples and morphological measurements.

Y.E.G. and P.G. interpreted the results and wrote the manuscript with the contribution of all other co-authors.

Data Accessibility

The nucleotide sequences were deposited in Genbank, under accession numbers KC431254 - KC431778. Microsatellite genotyping and morphological measurement for each individual will be deposited in the Dryad data repository. For further information on sample locations and microsatellite-based assignment scores, see Table S1, Supporting information.

Figure 1. Plots of the probabilistic assignment inferred in STRUCTURE for simulated (above) and observed (below) genotypes. Each specimen is represented by a vertical bar fragmented in *K* (= 2) colored sections that are relative to their membership proportion into one of the two genetic clusters, *O. agassizii* (dark grey) and *O. luteus* (light grey). The box delimitates the estimated threshold of *Tqi* = 0.86 in the assignment of simulated genotypes using HYBRIDLAB.

Figure 2. Morphotypes of pure parental specimens of *O. agassizii* and *O. luteus* and their hybrids identified from our genetic analyses in Lake Titicaca (A) and Lake Uru Uru (B).

Figure 3. Pattern of genetic differentiation among *O. agassizii*, *O. luteus* and hybrids based on factorial correspondence analysis (FCA) of individual multilocus genotypes, plotted on the two main contributing axes. Dark grey squares: *O. agassizii*; light grey squares: *O. luteus;* white squares: hybrids.

Figure 4. Estimated allele frequencies for the 10 microsatellites used in this study, between pure parental specimens of *O. agassizii* (black) and *O. luteus* (white).

Figure 5. Plot of canonical discriminant function based on 17 morphological variables for *O. agassizii* (squares), *O. luteus* (diamonds) and their hybrids (circles), for males (A) and females (B). Centroid for each group is given.

Table 1. Individual assignment (STRUCTURE) and inferred ancestry (NEWHYBRIDS) for *O. agassizii*, *O. luteus* and their hybrids. The most likely classes for hybrids appear in bold. Probability intervals for the two pure parental species are given in parentheses.

	Individual assignment (STRUCTURE)		Inferred ancestry (NEWHYBRIDS)					
	O. agassizii	*O. luteus*	*O. agassizii*	*O. luteus*	F1	F2	BxOa	BxOl
O. agassizii	0.993 (0.953–0.996)	0.007 (0.004–0.028)	0.998 (0.984–0.999)	-				
O. luteus	0.008 (0.003–0.023)	0.992 (0.972–0.996)	-	0.996 (0.987–0.998)				
GENT5	0.417	0.583	0.000	0.002	**0.543**	0.227	0.015	0.212
GENT10	0.333	0.667	0.000	0.007	**0.601**	0.113	0.003	0.276
GENT12	0.497	0.503	0.000	0.001	**0.957**	0.013	0.006	0.022
GENT14	0.282	0.718	0.000	0.008	**0.883**	0.017	0.002	0.091
GENT16	0.474	0.526	0.000	0.000	**0.934**	0.025	0.012	0.028
GENT18	0.424	0.576	0.000	0.001	**0.925**	0.021	0.005	0.047
U167	0.529	0.471	0.000	0.000	**0.968**	0.012	0.009	0.011
U170	0.428	0.572	0.000	0.000	**0.952**	0.014	0.004	0.029
U188	0.562	0.438	0.000	0.002	**0.876**	0.048	0.019	0.055
U197	0.526	0.474	0.000	0.002	**0.656**	0.163	0.008	0.172
U198	0.253	0.747	0.000	0.006	**0.755**	0.051	0.002	0.186
U205	0.733	0.267	0.005	0.000	**0.933**	0.022	0.034	0.007
U216	0.556	0.444	0.000	0.000	**0.946**	0.022	0.018	0.014
2U384	0.578	0.422	0.000	0.005	**0.904**	0.035	0.015	0.041
2U383	0.359	0.641	0.000	0.001	**0.882**	0.035	0.005	0.077
2U387	0.293	0.707	0.000	0.009	**0.551**	0.136	0.004	0.299
2U394	0.119	0.881	0.000	**0.434**	0.082	0.081	0.001	**0.402**

Table 2. Mitochondrial DNA (control region and cytochrome *b*) and nuclear DNA (rhodopsin) sequence diversity. Number of haplotypes (*h*), haplotype diversity (*Hd*), number of polymorphic sites (*S*), nucleotide diversity (*Pi*), number of sequences (N) and number of nucleotide for each locus (L).

Marker	Summary statistics	*O. agassizii*	*O. luteus*
Control region (mtDNA)	N	108	50
	L	397–8	398
	h	22	18
	Hd	0.701 (SD ± 0.046)	0.88 (SD ± 0.033)
	S	18	14
	Pi	0.003	0.005
	Overall mean distance between haplotypes	1.32	2.34
Cytochrome *b* (mtDNA)	N	108	50
	L	1102	1102
	h	51	14
	Hd	0.958 (SD ± 0.008)	0.627 (SD ± 0.075)
	S	59	15
	Pi	0.003	0.0009
	Overall mean distance between haplotypes	4.22	1.02
Rhodopsin (nDNA)	N	108	50
	L	689	689
	h	3	2
	Hd	0.048 (SD ± 0.001)	0.078 (SD ± 0.002)
	S	3	1
	Pi	0.000	0.000
	Overall mean distance between haplotypes	0.073	0.078

Table 3. Diagnostic sites in rhodopsin nucleotide sequences discriminating between *O. agassizii* and *O. luteus*, showing heterozygosity in hybrids. * Numbers refer to the position of each site according to our sequence alignment.

	Nucleotide position*						
	89	95	227	284	426	547	548
O. agassizii	C	C	G	C	C	T	G
O. luteus	T	T	A	T	G	G	C
Hybrids	Y	Y	R	Y	S	K	S

Table 4. Microsatellites diversity in *O. agassizii* (*O. aga*) and *O. luteus* (*O. lut*). Abbreviations: number of alleles (*NA*), number of private alleles (*NPA*), allelic richness (*AR*), private allelic richness (*PAR*) estimated after rarefaction, observed and expected heterozygosities (H_O and H_E, respectively), and exact *P*-values for HWE test.

Locus	Size range		*NA*		*NPA*		*AR*		*PAR*		H_O		H_E		HWE *P*-value	
	O. aga	*O. lut*	*O. aga*	*O. lut*	*O. aga*	*O. lut*	*O. aga*	*O. lut*	*O. aga*	*O. lut*	*O. aga*	*O. lut*	*O. aga*	*O. lut*	*O. aga*	*O. lut*
A9a	228–270	234–240	17	3	17	3	13.92	2.94	13.92	2.94	0.84	0.12	0.88	0.11	0.06	1
A106	254–320	248–322	31	30	6	5	25.08	29.85	3.58	8.35	0.88	0.94	0.95	0.96	0.03	0.84
A116	300–384	288–300	34	7	33	6	26.13	6.99	25.75	6.61	0.78	0.57	0.92	0.79	0.04	0 *
B1	169–183	171–193	9	6	4	1	7.28	5.52	3.00	1.24	0.38	0.44	0.42	0.52	0.03	0.17
B103	185–223	195–223	9	7	5	3	5.98	6.63	2.80	3.45	0.59	0.76	0.59	0.71	0.96	0.51
B104	136–189	141–183	18	2	18	2	14.28	1.77	14.28	1.77	0.81	0.20	0.78	0.20	0.21	-
C101	212–318	254–318	28	17	14	3	22.13	16.70	11.73	6.31	0.81	0.91	0.92	0.93	0.21	0.11
C102	225–237	239–281	5	12	5	12	3.93	11.74	3.93	11.74	0.14	0.88	0.16	0.87	0.02	0.39
C105	192–254	196–258	16	18	5	7	13.01	17.11	3.54	7.65	0.87	0.75	0.86	0.92	0.20	0.05
D110	232–350	244–322	26	20	8	2	21.27	18.74	6.13	3.60	0.86	0.94	0.94	0.92	0.22	0.24
Mean	-	-	19.3	12.2	11.5	4.4	15.30	11.79	8.86	5.36	0.69	0.65	0.74	0.69	-	-

* Significant departure from HWE after Bonferroni correction ($P < 0.006$).

Table 5. Structure matrix of discriminant loadings from discriminant analysis (DA) based on 17 morphometric variables. Coefficients above 0.50 or below -0.50 are highlighted in bold.

Morphometric	Male		Female	
	CF1	CF2	CF1	CF2
Head width	**-0.669**	-0.228	**-0.679**	-0.067
Caudal fin length	0.450	0.214	0.295	0.291
Head length	-0.374	-0.201	-0.335	0.066
Pectoral fin length	0.368	0.124	0.093	-0.176
Standard length	0.336	0.093	0.344	0.378
Mouth height	-0.317	-0.004	-0.331	-0.075
Pectoral fin height	-0.159	-0.021	-0.222	-0.121
Maximum body height	-0.153	-0.138	-0.024	0.023
Distance between eyes	-0.115	-0.010	-0.054	-0.110
Body width	-0.232	**-0.502**	-0.171	-0.142
Eye diameter	0.003	0.434	-0.027	0.324
Caudal peduncle length	0.319	0.384	0.324	0.158
Caudal peduncle height	0.119	-0.376	0.162	-0.288
Eye height	-0.110	0.185	-0.061	0.234
Caudal peduncle width	0.132	-0.157	0.221	0.159
Head height	0.116	-0.136	0.062	-0.342
Caudal fin height	-0.024	-0.038	-0.002	0.008

Table 6. Cross-validated classification matrix for *O. agassizii*, *O. luteus* and their hybrids based on a discriminant analysis of 17 morphometric variables. Males are shown above and females below.

Original group	Predicted group membership (%)			Total counts
	O. agassizii	*O. luteus*	Hybrids	
O. agassizii	15 (71.4%) 81 (93.1%)	1 (4.8%) 0 (0%)	5 (23.8%) 6 (6.9%)	21 87
O. luteus	0 (-) 0 (-)	33 (100%) 17 (100%)	0 (-) 0 (-)	33 17
Hybrids	3 (50%) 1 (9.1%)	0 (-) 1 (9.1%)	3 (50%) 9 (81.8%)	6 11

Table 7. Assessment of Lauzanne's (1982) morphometric criteria to discriminate between *O. agassizii* and *O. luteus*. Ratios are expressed in percentage; the mean and standard deviation are specified in parentheses. Standard length (SL) is given in millimeters. Morphometric criteria are given as a ratio with the standard length (A).

Morphometric criteria (ratios)	**Lauzanne's study**		**This study**		
	O. agassizii SL = 30.5–151.0	*O. luteus* SL = 68.0–105.0	*O. agassizii* SL = 78.45–150.2	*O. luteus* SL = 76.39–118.8	Hybrids SL = 47.92–138.77
Head length (B/A)	24–28% (25.9 ± 0.9)	31–36% (34.0 ± 1.16)	20–30% (25.15 ± 2.27)	26–38% (34.20 ± 2.44)	25–58% (30.19 ± 7.58)
Head width (N/A)	13-24 % (18.2 ± 2.57)	25-34% (29.6 ± 2.33)	11-23% (17.26 ± 2.22)	20-36% (29.35 ± 3.20)	17-41% (22.79 ± 5.36)
Body height (F/A)	25-32% (28.4 ± 1.5)	33-40% (35.7 ± 1.67)	19-36% (29.26 ± 3.46)	26-46% (36.57 ± 3.49)	21-69% (34.27 ± 10.13)

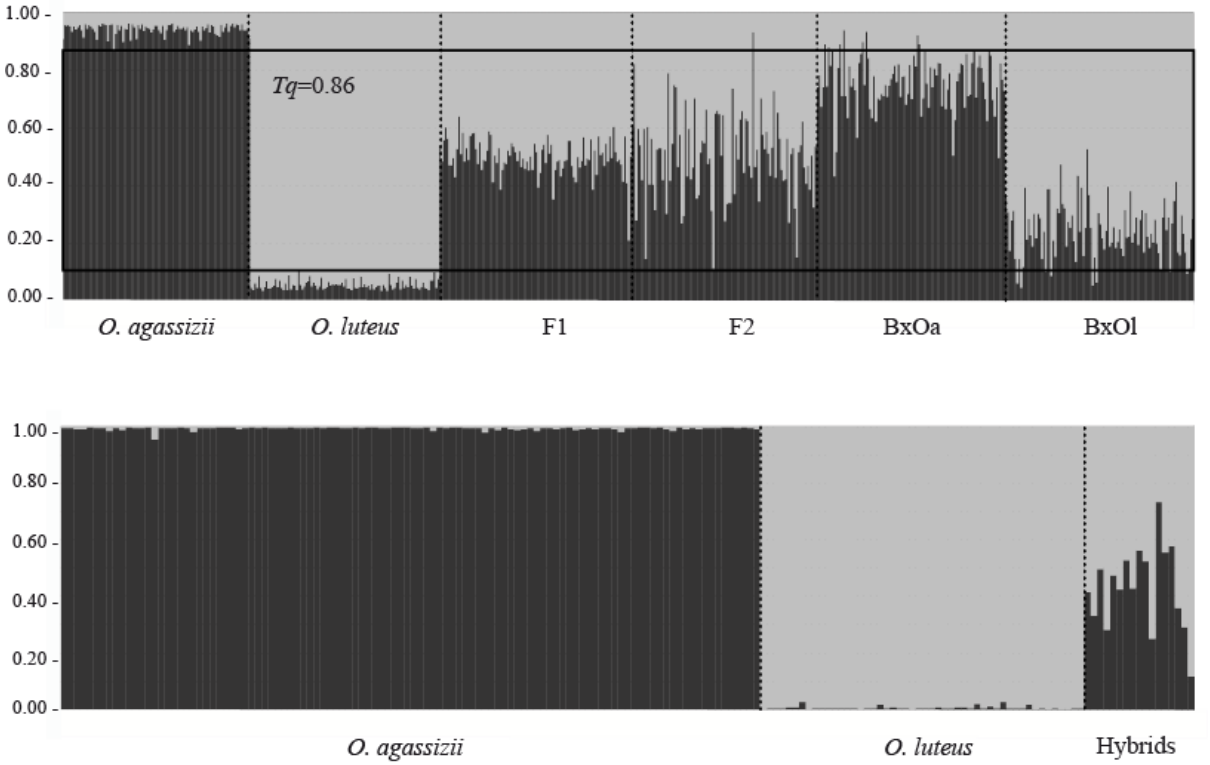

1.00
0.80
0.60
0.40
0.20
0.00
Tq=0.86
O. agassizii
O. luteus
F1
F2
BxOa
BxOl
1.00
0.80
0.60
0.40
0.20
0.00
O. agassizii
O. luteus
Hybrids

Figure 1

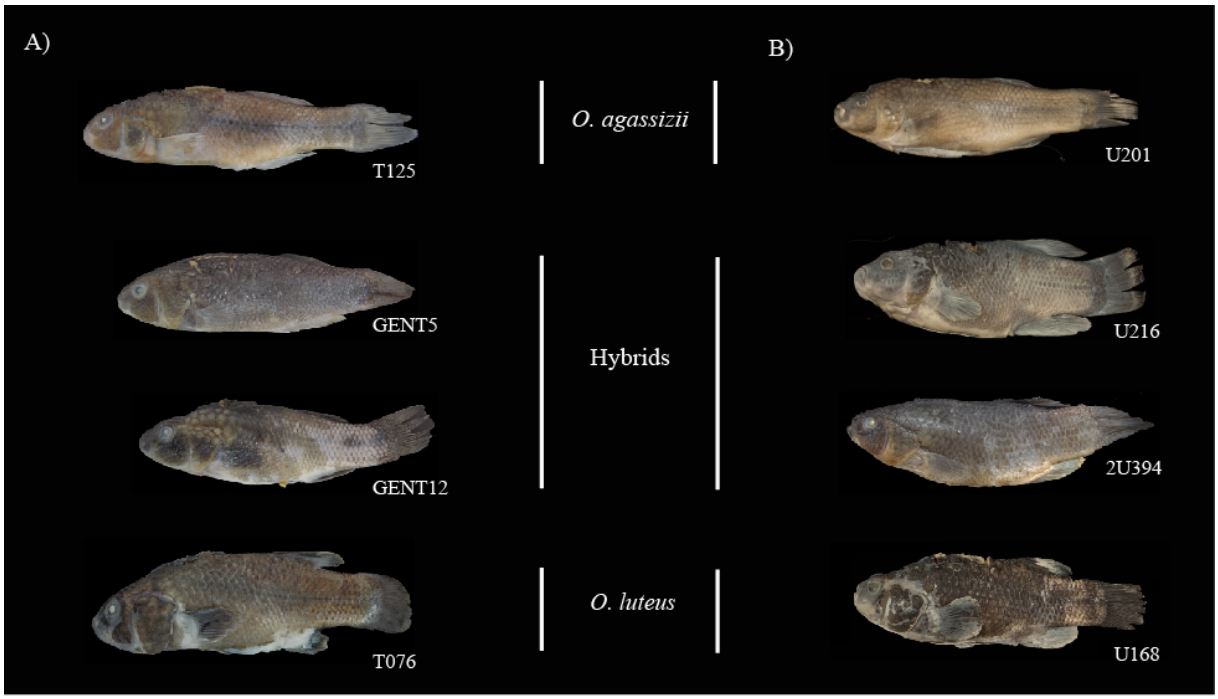

A)
B)
T125
O. agassizii
U201
GENT5
Hybrids
U216
GENT12
2U394
T076
O. luteus
U168

Figure 2

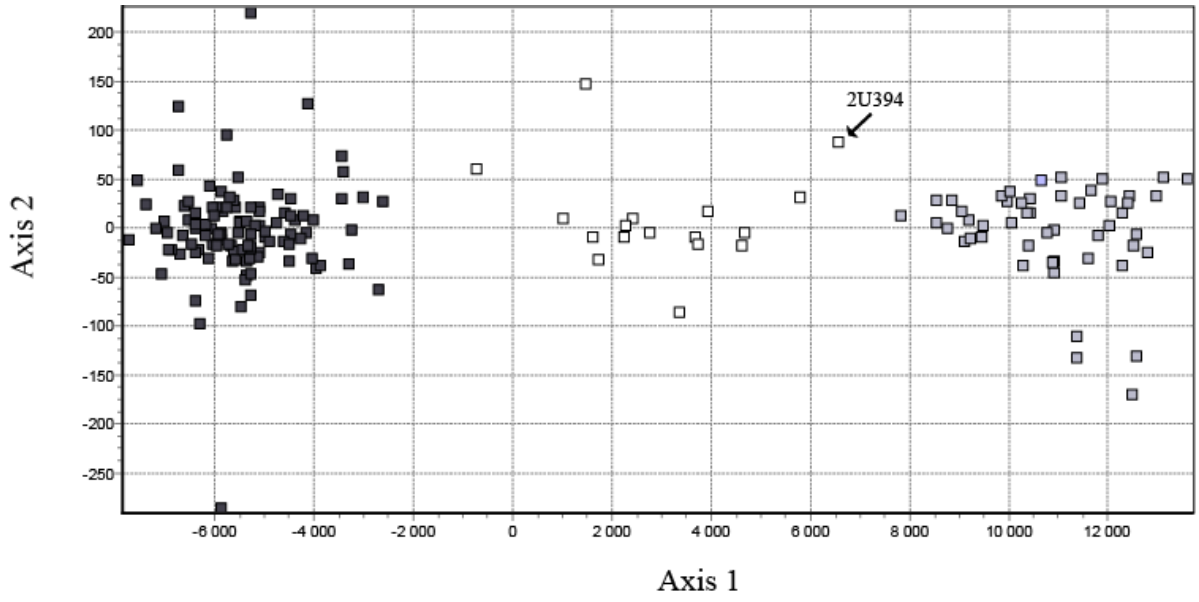

2U394
Axis 2
Axis 1
200
150
100
50
0
-50
-100
-150
-200
-250
-6 000
-4 000
-2 000
0
2 000
4 000
6 000
8 000
10 000
12 000

Figure 3

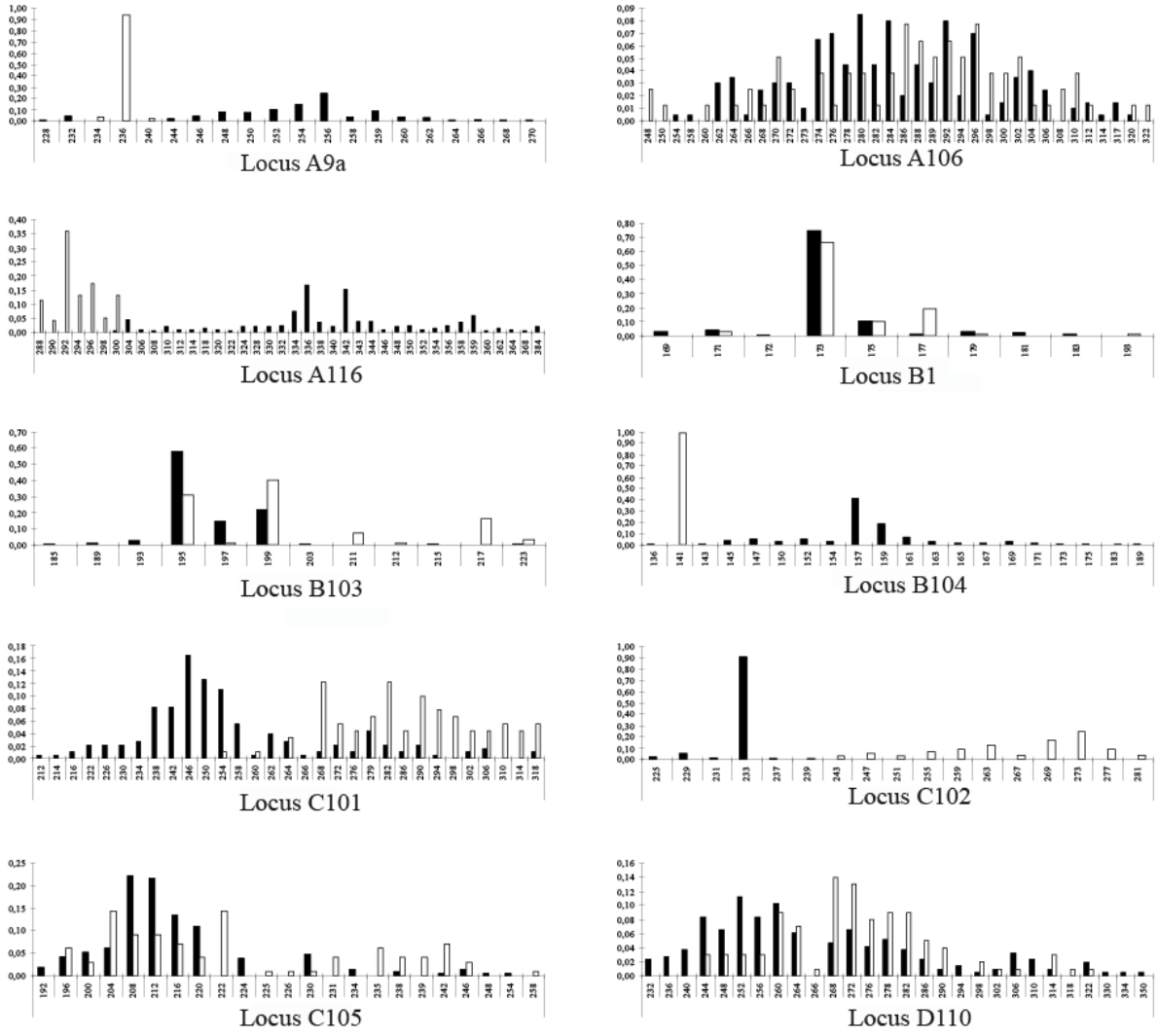
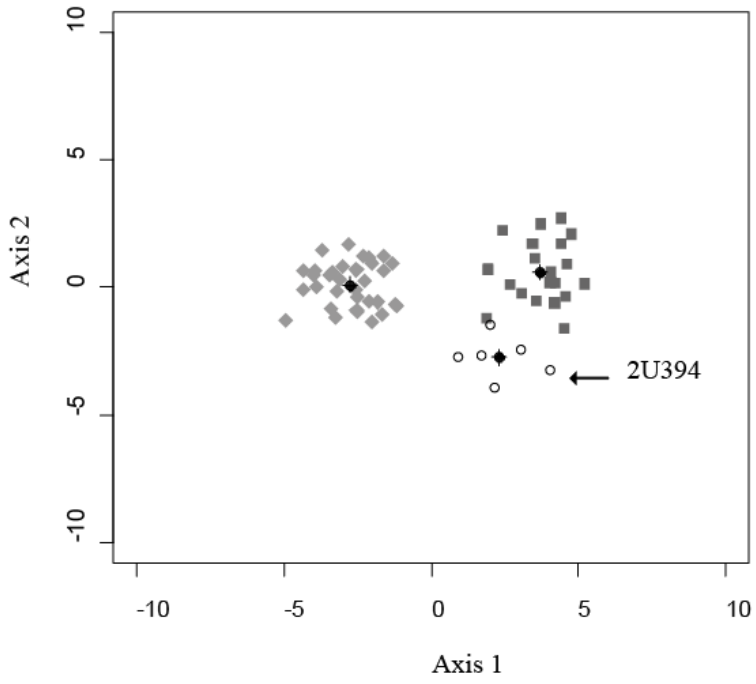

Locus A9a
Locus A106
Locus A116
Locus B1
Locus B103
Locus B104
Locus C101
Locus C102
Locus C105
Locus D110

Figure 4

A)


Axis 2
10
5
0
-5
-10
-10
-5
0
5
10
Axis 1
2U394

B)

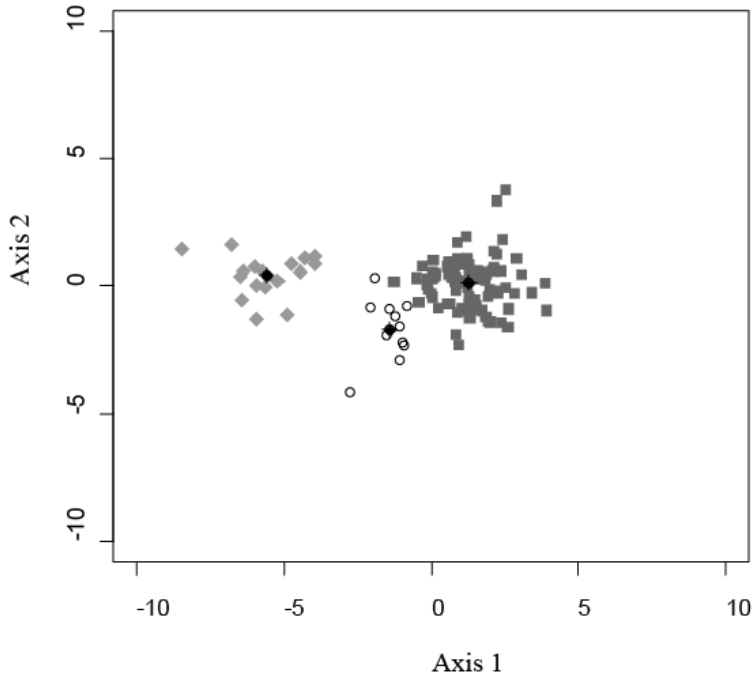

Axis 2
10
5
0
-5
-10
-10
-5
0
5
10
Axis 1

Figure 5

Table S1. Genetic identity (ID) of the specimens under study, including their inferred genotype frequency classes in STRUCTURE (left; color code: red = *O. agassizii*, blue = *O. luteus*, purple = hybrids) and NEWHYBRIDS (right) using microsatellites (see Table 1 for hybrids), their lineage identity based on rhodopsin (RH), control region (CR) and cytochrome *b* (Cyt*b*) nucleotide sequences, and their posterior classification probabilities based on cross-validation using discriminant analysis (DA) of 17 standard morphometric characters. For RH, specimens showing heterozygosity at non-diagnostic sites are in light orange and those showing heterozygosity at all the diagnostic synapomorphic sites are in black (see Results). * Meristic criteria from Lauzanne (1982). Abbreviations: *O. aga* = *O. agassizii*; *O. lut* = *O. luteus*; NG = non-granular scales; G = granular scales; GS = Gonadal state (from 1-juvenile to 5-running stage).

ID	Sample code	Microsatellites		RH	CR	Cyt*b*	Posterior classification probabilities (DA)			Type of scales on the head*	Sample locality	Sex	GS
							O. aga	*O. lut*	hybrids				
O. agassizii	T095	0.995	0.999				0.253	0.000	0.747	NG	Huatajata, Lake Titicaca	F	5
O. agassizii	T099	0.994	0.998				0.531	0.000	0.469	NG	Huatajata, Lake Titicaca	F	5
O. agassizii	T100	0.991	0.999				0.999	0.000	0.001	NG	Huatajata, Lake Titicaca	M	3
O. agassizii	T101	0.992	0.997				1.000	0.000	0.000	NG	Huatajata, Lake Titicaca	M	3
O. agassizii	T103	0.996	1.000				1.000	0.000	0.000	NG	Huatajata, Lake Titicaca	M	2
O. agassizii	T104	0.995	1.000				1.000	0.000	0.000	NG	Huatajata, Lake Titicaca	M	3
O. agassizii	T105	0.995	0.999				0.000	0.955	0.032	NG	Huatajata, Lake Titicaca	M	2
O. agassizii	T106	0.983	0.997				0.452	0.000	0.300	NG	Huatajata, Lake Titicaca	M	2
O. agassizii	T107	0.993	0.999				0.957	0.000	0.043	NG	Huatajata, Lake Titicaca	M	3
O. agassizii	T109	0.988	0.996				0.338	0.000	0.662	NG	Huatajata, Lake Titicaca	M	2
O. agassizii	T110	0.996	0.999				0.916	0.000	0.084	NG	Huatajata, Lake Titicaca	M	2
O. agassizii	T111	0.995	1.000				0.985	0.000	0.015	NG	Huatajata, Lake Titicaca	M	2
O. agassizii	T113	0.995	0.999				1.000	0.000	0.000	NG	Huatajata, Lake Titicaca	M	2
O. agassizii	T114	0.996	1.000				0.640	0.000	0.360	NG	Huatajata, Lake Titicaca	F	2

O. agassizii	T115	0.953	0.984				0.271	0.000	0.729	NG	Huatajata, Lake Titicaca	M	2
O. agassizii	T116	0.995	1.000				1.000	0.000	0.000	NG	Huatajata, Lake Titicaca	M	2
O. agassizii	T117	0.994	0.999				1.000	0.000	0.000	NG	Huatajata, Lake Titicaca	M	2
O. agassizii	T118	0.993	0.998				0.982	0.000	0.018	NG	Huatajata, Lake Titicaca	F	3
O. agassizii	T119	0.996	1.000				1.000	0.000	0.000	NG	Huatajata, Lake Titicaca	F	5
O. agassizii	T120	0.994	0.999				0.999	0.000	0.001	NG	Huatajata, Lake Titicaca	F	3
O. agassizii	T122	0.980	0.993				1.000	0.000	0.000	NG	Huatajata, Lake Titicaca	F	5
O. agassizii	T123	0.994	0.999				0.442	0.000	0.558	NG	Huatajata, Lake Titicaca	M	2
O. agassizii	T124	0.993	0.998				0.999	0.000	0.001	NG	Huatajata, Lake Titicaca	M	1
O. agassizii	T125	0.993	0.998				0.997	0.000	0.003	NG	Huatajata, Lake Titicaca	F	2
O. agassizii	T126	0.996	1.000				0.048	0.000	0.952	NG	Huatajata, Lake Titicaca	M	2
O. agassizii	T127	0.996	1.000				0.982	0.000	0.018	NG	Huatajata, Lake Titicaca	F	3
O. agassizii	T128	0.996	1.000				1.000	0.000	0.000	NG	Huatajata, Lake Titicaca	F	3
O. agassizii	T129	0.991	0.999				0.974	0.000	0.026	NG	Huatajata, Lake Titicaca	F	3
O. agassizii	T130	0.995	0.999				0.731	0.000	0.269	NG	Huatajata, Lake Titicaca	F	5
O. agassizii	T131	0.996	1.000				0.999	0.000	0.001	NG	Huatajata, Lake Titicaca	M	2
O. agassizii	T132	0.995	0.999				1.000	0.000	0.000	NG	Huatajata, Lake Titicaca	M	2
O. agassizii	T133	0.996	1.000				1.000	0.000	0.000	NG	Huatajata, Lake Titicaca	F	2
O. agassizii	T134	0.995	1.000				0.980	0.000	0.020	NG	Huatajata, Lake Titicaca	F	4
O. agassizii	T135	0.995	0.999				1.000	0.000	0.000	NG	Huatajata, Lake Titicaca	F	3
O. agassizii	T136	0.994	0.999				1.000	0.000	0.000	NG	Huatajata, Lake Titicaca	F	2
O. agassizii	T137	0.996	1.000				0.950	0.000	0.050	NG	Huatajata, Lake Titicaca	F	1
O. agassizii	T140	0.995	1.000				0.989	0.000	0.011	NG	Huatajata, Lake Titicaca	F	1
O. agassizii	GENT3	0.993	0.999				0.999	0.000	0.001	NG	Huatajata, Lake Titicaca	F	3
O. agassizii	GENT4	0.995	1.000				1.000	0.000	0.000	NG	Huatajata, Lake Titicaca	F	5
O. agassizii	GENT29	0.996	1.000				0.998	0.000	0.002	NG	Huatajata, Lake Titicaca	F	5
O. agassizii	GENT33	0.995	1.000				1.000	0.000	0.000	NG	Huatajata, Lake Titicaca	F	5
O. agassizii	U187	0.996	1.000				0.918	0.000	0.082	NG	Lake Uru Uru	F	5
O. agassizii	U189	0.996	1.000				0.384	0.000	0.616	NG	Lake Uru Uru	F	5
O. agassizii	U191	0.993	0.999				0.897	0.000	0.103	NG	Lake Uru Uru	F	5

O. agassizii	U192	0.995	0.999				0.556	0.000	0.444	NG	Lake Uru Uru	F	5
O. agassizii	U193	0.996	1.000				0.210	0.000	0.790	NG	Lake Uru Uru	F	4
O. agassizii	U195	0.993	0.998				0.794	0.000	0.206	NG	Lake Uru Uru	F	3
O. agassizii	U196	0.996	1.000				0.988	0.000	0.012	NG	Lake Uru Uru	F	3
O. agassizii	U199	0.996	1.000				0.997	0.000	0.003	NG	Lake Uru Uru	F	5
O. agassizii	U200	0.994	0.999				0.036	0.000	0.963	NG	Lake Uru Uru	F	2
O. agassizii	U201	0.993	0.997				0.248	0.000	0.752	NG	Lake Uru Uru	F	4
O. agassizii	U202	0.996	0.999				0.913	0.000	0.087	NG	Lake Uru Uru	F	2
O. agassizii	U203	0.996	1.000				0.005	0.000	0.995	NG	Lake Uru Uru	M	2
O. agassizii	U204	0.997	1.000				0.641	0.000	0.359	NG	Lake Uru Uru	M	1
O. agassizii	U206	0.995	0.999				1.000	0.000	0.000	NG	Lake Uru Uru	F	3
O. agassizii	U207	0.995	0.999				0.946	0.000	0.054	NG	Lake Uru Uru	F	4
O. agassizii	U208	0.996	1.000				0.621	0.000	0.379	NG	Lake Uru Uru	F	3
O. agassizii	U209	0.983	0.998				0.999	0.000	0.001	NG	Lake Uru Uru	F	2
O. agassizii	U210	0.996	1.000				0.986	0.000	0.014	NG	Lake Uru Uru	F	4
O. agassizii	U211	0.994	0.999				0.873	0.000	0.127	NG	Lake Uru Uru	F	3
O. agassizii	U212	0.997	1.000				0.946	0.000	0.054	NG	Lake Uru Uru	F	5
O. agassizii	U213	0.996	1.000				0.999	0.000	0.001	NG	Lake Uru Uru	F	5
O. agassizii	U214	0.994	0.998				0.988	0.000	0.012	NG	Lake Uru Uru	F	5
O. agassizii	U215	0.993	0.999				0.990	0.000	0.010	NG	Lake Uru Uru	F	5
O. agassizii	1U375	0.995	0.999				0.860	0.000	0.140	NG	Lake Uru Uru	F	2
O. agassizii	1U376	0.977	0.997				0.756	0.000	0.244	NG	Lake Uru Uru	F	2
O. agassizii	1U377	0.994	0.999				1.000	0.000	0.000	NG	Lake Uru Uru	F	2
O. agassizii	1U378	0.988	0.995				1.000	0.000	0.000	NG	Lake Uru Uru	F	1
O. agassizii	1U379	0.996	1.000				0.919	0.000	0.081	NG	Lake Uru Uru	F	2
O. agassizii	1U380	0.992	0.998				0.985	0.000	0.015	NG	Lake Uru Uru	F	2
O. agassizii	1U381	0.989	0.997				0.998	0.000	0.002	NG	Lake Uru Uru	F	3
O. agassizii	2U382	0.990	0.996				0.789	0.000	0.211	NG	Lake Uru Uru	F	1
O. agassizii	2U385	0.994	0.999				0.959	0.000	0.041	NG	Lake Uru Uru	F	3
O. agassizii	2U386	0.984	0.997				0.997	0.000	0.003	NG	Lake Uru Uru	F	2

O. agassizii	2U388	0.994	0.999				0.989	0.000	0.011	NG	Lake Uru Uru	F	3
O. agassizii	2U389	0.995	0.999				0.993	0.000	0.007	NG	Lake Uru Uru	F	2
O. agassizii	2U390	0.992	0.998				0.999	0.000	0.001	NG	Lake Uru Uru	F	4
O. agassizii	2U391	0.993	0.999				0.990	0.000	0.010	NG	Lake Uru Uru	F	2
O. agassizii	2U392	0.996	1.000				0.972	0.000	0.028	NG	Lake Uru Uru	F	4
O. agassizii	2U393	0.987	0.996				0.999	0.000	0.001	NG	Lake Uru Uru	F	2
O. agassizii	2U395	0.992	0.998				0.980	0.000	0.020	NG	Lake Uru Uru	F	3
O. agassizii	P226	0.995	0.999				0.983	0.000	0.017	NG	Lake Poopó	F	3
O. agassizii	P227	0.992	0.997				0.989	0.000	0.011	NG	Lake Poopó	F	4
O. agassizii	P228	0.993	0.999				0.998	0.000	0.002	NG	Lake Poopó	F	5
O. agassizii	P229	0.995	0.999				0.999	0.000	0.001	NG	Lake Poopó	F	5
O. agassizii	P230	0.992	0.998				1.000	0.000	0.000	NG	Lake Poopó	F	5
O. agassizii	P231	0.981	0.989				0.998	0.000	0.002	NG	Lake Poopó	F	3
O. agassizii	P233	0.995	0.999				0.990	0.000	0.010	NG	Lake Poopó	F	2
O. agassizii	P234	0.995	0.999				0.997	0.000	0.003	NG	Lake Poopó	F	5
O. agassizii	P235	0.996	1.000				0.921	0.000	0.079	NG	Lake Poopó	F	2
O. agassizii	P236	0.996	0.999				0.986	0.000	0.014	NG	Lake Poopó	F	3
O. agassizii	P237	0.995	0.999				0.998	0.000	0.002	NG	Lake Poopó	F	2
O. agassizii	P238	0.997	1.000				1.000	0.000	0.000	NG	Lake Poopó	F	3
O. agassizii	P239	0.990	0.998				0.995	0.000	0.005	NG	Lake Poopó	F	2
O. agassizii	P241	0.986	0.996				0.983	0.000	0.017	NG	Lake Poopó	F	4
O. agassizii	P242	0.996	0.999				0.999	0.000	0.001	NG	Lake Poopó	F	2
O. agassizii	P243	0.991	0.996				0.984	0.000	0.016	NG	Lake Poopó	F	5
O. agassizii	P244	0.993	0.999				0.999	0.000	0.001	NG	Lake Poopó	F	3
O. agassizii	P245	0.992	0.998				0.065	0.000	0.935	NG	Lake Poopó	F	2
O. agassizii	P246	0.995	0.999				1.000	0.000	0.000	NG	Lake Poopó	F	3
O. agassizii	P247	0.995	0.999				0.924	0.000	0.076	NG	Lake Poopó	F	2
O. agassizii	P248	0.995	0.999				0.999	0.000	0.001	NG	Lake Poopó	F	5
O. agassizii	P249	0.996	1.000				0.970	0.000	0.030	NG	Lake Poopó	F	4
O. agassizii	P250	0.996	1.000				1.000	0.000	0.000	NG	Lake Poopó	F	4

O. agassizii	P251	0.995	0.999				0.813	0.000	0.187	NG	Lake Poopó	F	3
O. agassizii	P252	0.996	1.000				1.000	0.000	0.000	NG	Lake Poopó	F	2
O. agassizii	P253	0.995	0.999				1.000	0.000	0.000	NG	Lake Poopó	F	2
O. agassizii	P254	0.992	0.999				1.000	0.000	0.000	NG	Lake Poopó	F	5
O. luteus	GENT6	0.004	0.998				0.000	0.985	0.015	G	Huatajata, Lake Titicaca	F	5
O. luteus	GENT7	0.005	0.997				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	F	5
O. luteus	GENT8	0.006	0.996				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	M	2
O. luteus	GENT9	0.005	0.997				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	F	3
O. luteus	GENT11	0.008	0.997				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	F	5
O. luteus	GENT13	0.008	0.994				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	M	2
O. luteus	GENT15	0.028	0.994				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	M	2
O. luteus	GENT17	0.004	0.998				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	M	3
O. luteus	GENT20	0.005	0.999				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	M	2
O. luteus	GENT21	0.005	0.998				0.000	0.952	0.048	G	Huatajata, Lake Titicaca	F	3
O. luteus	GENT22	0.005	0.997				0.000	0.999	0.000	G	Huatajata, Lake Titicaca	M	2
O. luteus	GENT23	0.007	0.995				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	M	2
O. luteus	GENT24	0.005	0.998				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	M	2
O. luteus	GENT26	0.007	0.998				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	F	2
O. luteus	GENT27	0.006	0.995				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	M	2
O. luteus	GENT28	0.004	0.999				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	M	2
O. luteus	GENT30	0.005	0.999				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	M	2
O. luteus	GENT32	0.005	0.997				0.000	0.992	0.008	G	Huatajata, Lake Titicaca	F	5
O. luteus	T075	0.019	0.992				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	F	5
O. luteus	T076	0.005	0.998				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	M	2
O. luteus	T077	0.009	0.997				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	F	3
O. luteus	T078	0.007	0.995				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	F	2
O. luteus	T079	0.006	0.994				0.000	1.000	0.000	G	Huatajata, Lake Titicaca	M	2
O. luteus	T080	0.004	0.998				0.000	0.997	0.003	G	Huatajata, Lake Titicaca	F	5
O. luteus	T096	0.005	0.998				0.000	0.992	0.008	G	Huatajata, Lake Titicaca	F	3
O. luteus	U166	0.007	0.996				0.000	1.000	0.000	G	Lake Uru Uru	F	5

O. luteus	U168	0.005	0.996				0.000	1.000	0.000	G	Lake Uru Uru	M	2
O. luteus	U169	0.010	0.997				0.000	1.000	0.000	G	Lake Uru Uru	F	5
O. luteus	U171	0.006	0.997				0.000	1.000	0.000	G	Lake Uru Uru	M	2
O. luteus	U172	0.005	0.998				0.000	1.000	0.000	G	Lake Uru Uru	M	2
O. luteus	U173	0.009	0.996				0.000	1.000	0.000	G	Lake Uru Uru	M	2
O. luteus	U174	0.010	0.997				0.000	1.000	0.000	G	Lake Uru Uru	M	2
O. luteus	U175	0.005	0.997				0.000	0.999	0.001	G	Lake Uru Uru	F	5
O. luteus	U176	0.023	0.989				0.000	1.000	0.000	G	Lake Uru Uru	M	2
O. luteus	U177	0.005	0.997				0.000	1.000	0.000	G	Lake Uru Uru	M	2
O. luteus	U178	0.013	0.994				0.000	1.000	0.000	G	Lake Uru Uru	M	2
O. luteus	U180	0.005	0.999				0.000	1.000	0.000	G	Lake Uru Uru	M	1
O. luteus	U182	0.028	0.987				0.000	1.000	0.000	G	Lake Uru Uru	M	1
O. luteus	U183	0.006	0.995				0.000	0.780	0.218	G	Lake Uru Uru	M	1
O. luteus	U184	0.006	0.997				0.000	0.995	0.005	G	Lake Uru Uru	M	2
O. luteus	U185	0.005	0.997				0.000	1.000	0.000	G	Lake Uru Uru	M	1
O. luteus	U186	0.019	0.989				0.000	1.000	0.000	G	Lake Uru Uru	F	2
O. luteus	U217	0.004	0.998				0.000	0.904	0.095	G	Lake Uru Uru	M	2
O. luteus	U218	0.007	0.994				0.000	1.000	0.000	G	Lake Uru Uru	M	1
O. luteus	U219	0.004	0.999				0.000	1.000	0.000	G	Lake Uru Uru	M	2
O. luteus	U220	0.005	0.998				0.000	0.975	0.025	G	Lake Uru Uru	M	2
O. luteus	U221	0.004	0.999				0.000	0.993	0.007	G	Lake Uru Uru	F	5
O. luteus	U222	0.004	0.999				0.000	1.000	0.000	G	Lake Uru Uru	M	5
O. luteus	U223	0.005	0.998				0.000	1.000	0.000	G	Lake Uru Uru	M	1
O. luteus	U224	0.006	0.995				0.000	1.000	0.000	G	Lake Uru Uru	M	2
hybrids	GENT5						0.990	0.000	0.010	NG	Huatajata, Lake Titicaca	M	2
hybrids	GENT10						0.339	0.000	0.661	NG	Huatajata, Lake Titicaca	F	3
hybrids	GENT12						0.244	0.000	0.753	NG	Huatajata, Lake Titicaca	M	2
hybrids	GENT14						0.269	0.000	0.637	NG	Huatajata, Lake Titicaca	F	3
hybrids	GENT16						0.169	0.000	0.831	NG	Huatajata, Lake Titicaca	F	2
hybrids	GENT18						0.999	0.000	0.001	NG	Huatajata, Lake Titicaca	M	2

hybrids	U167						0.008	0.000	0.992	NG	Lake Uru Uru	F	5
hybrids	U170						0.021	0.000	0.979	NG	Lake Uru Uru	F	5
hybrids	U188						0.295	0.000	0.704	NG	Lake Uru Uru	F	5
hybrids	U197						0.013	0.000	0.980	NG	Lake Uru Uru	F	5
hybrids	U198						0.014	0.000	0.986	NG	Lake Uru Uru	F	3
hybrids	U205						0.000	0.005	0.992	NG	Lake Uru Uru	M	1
hybrids	U216						0.016	0.000	0.984	G	Lake Uru Uru	M	3
hybrids	2U384						0.003	0.000	0.997	NG	Lake Uru Uru	F	1
hybrids	2U383						0.648	0.000	0.349	NG	Lake Uru Uru	F	1
hybrids	2U387						0.000	0.918	0.082	NG	Lake Uru Uru	F	3
hybrids	2U394						1.000	0.000	0.000	NG	Lake Uru Uru	M	1

II.5 Article 4

Yareli Esquer Garrigos, Bernard Hugueny, Carla Ibañez, Marco Méndez, Pamela Morales and Philippe Gaubert. **Phylogeographic analysis of the *agassizii* complex (genus *Orestias*: Cyprinodontidae): a rapid diversification following dispersal across the Andean Altiplano?** (in preparation).

Phylogeographic analysis of the *agassizii* complex (genus *Orestias*: Cyprinodontidae): a rapid diversification following dispersal across the Andean Altiplano?

In preparation

Yareli Esquer Garrigos [1], Bernard Hugueny [1], Carla Ibañez [1,2], Marco Méndez [3], Pamela Morales [3], Arnaud Couloux [4] and Philippe Gaubert [1]

[1] *UMR BOREA, Département Milieux et Peuplements Aquatiques, MNHN - CNRS 7208 - IRD 207 - UPMC, Muséum National d'Histoire Naturelle, 43 rue Cuvier, 75231 Paris, France*

[2] *Instituto de Ecología – Unidad de Limnología, Universidad Mayor de San Andrés, Casilla 10077, La Paz, Bolivia*

[3] *Departamento de Ciencias Ecológicas, Facultad de Ciencias, Universidad de Chile, Casilla 653, Santiago, Chile*

[4] *Genoscope, Centre National de Séquençage, 2 rue Gaston Crémieux, CP5706, 91057 Evry Cedex, France*

Corresponding author: Y. Esquer Garrigos. E-mail: esquer@mnhn.fr

Abstract

Diversification among freshwater taxa may occur via a combination of processes involving ecological speciation and dispersal regulated by the dynamics and habitat structure of lakes and stream networks. We conducted a phylogeographic analysis on an endemic complex of Andean pupfishes (*Orestias* '*agassizii*'), which is supposed to have undergone a rapid diversification from Lake Titicaca followed by dispersal across the Altiplano. Combining mitochondrial DNA and microsatellite loci, we tested the following hypotheses: (i) Lake

Titicaca was the center of dispersion of the *agassizii* complex, (ii) genetic connectivity among populations followed Pleistocene water level fluctuations, and (iii) geomorphologic distances shaped patterns of isolation in the complex. Our results support a 'source-sink' mode of diversification from Lake Titicaca, with a higher genetic diversity in the lake, a most widely distributed genetic group (microsatellites; group 1) —including the lake— across the Altiplano, and a positive correlation between geographic distances (altitude, Euclidian, hydrological) and genetic distances together with a negative correlation with genetic diversity, using the lake as a starting point. We also provided evidence for a major impact of paleolake fluctuations on population connectivity within the *agassizii* complex, notably through the superimposition of the boundaries of Paleolake Ballivian (middle Pleistocene) with the geographic delimitation of group 1. We also identified six isolated genetic populations corresponding to high altitude systems (Sur Lipez, Siete Lagunas, Ulla Ulla, Hichu Khota valley, and an isolated river (Rio Empexa) and a more widely distributed population showing some admixture with Lake Titicaca, both west of the Altiplano. Our analyses showed a strong effect of isolation-by-distance across the Altiplano, despite regular flooding and intermittent connections among basins and populations. Remarkably, elevation had the most significant effect on genetic distance and diversity distribution. We propose a scenario of recent diversification within the *agassizii* complex that combines ecological speciation within Lake Titicaca (diversification with inter-specific competition) and allopatric speciation outside the lake (diversification without inter-specific competition, following long-distance dispersal), the latter leading to genetically isolated populations restricted to fragile habitats that should be considered of conservation concern. Explaining how those Andean pupfishes with poor swimming abilities managed to disperse through a large altitudinal and ecological gradient across the Altiplano will require additional investigations on their life-history traits.

Keywords

Altiplano, *Orestias*, *agassizii* complex, paleolakes, Pleistocene, isolation-by-distance, genetic diversity, diversification

Introduction

In freshwater taxa, dispersal among drainages is dependent upon the connectivity of freshwater systems (Unmack 2001; Wong *et al.* 2004), which may change over geological processes through rearrangement of drainages and river courses (Wong *et al.* 2004). Over shorter temporal scales, floods may also connect waterways by inundating low divides (McGlashan & Hugues 2002; Wong *et al.* 2004). For organisms with high dispersal abilities, continuous range expansion or colonization of novel aquatic habitats may led to speciation through processes such as founder effect, natural selection and hybridization (Hughes *et al.* 2009; Cristescu *et al.* 2012).

Gene flow is generally expected to prevent or delay speciation by breaking linkage between genes involved in local adaptation and reproductive isolation (Coyne & Orr 2004; Kocher 2004; Matschiner *et al.* 2010). Thus, the identification of barriers to migration imposed by landscape features and their impact on population connectivity, genetic structure and differentiation is of central interest in population biology and conservation genetics (Castric *et al.* 2001; Costello *et al.* 2003). Nevertheless, superimposed processes including isolation-by-distance and ecological speciation may also promote speciation while gene flow still occurs (Kocher 2004; Seehausen 2004; Rundle & Nosil 2005; Garant *et al.* 2007) .

Patterns of dispersal and genetic structure among populations of freshwater taxa may be affected by extrinsic factors such as the structure of stream network, habitat discontinuity and water level fluctuations (Eckert *et al.*) (Koblmüller *et al.* 2007; Hughes *et al.* 2009; Matschiner *et al.* 2010). Distance may also affect connectivity, since populations separated by longer stretches of stream channel are likely to be more isolated (Hughes *et al.* 2009).

Elevation is a also factor of isolation, especially in widespread taxa that occur at high altitudes, where populations are more likely to be genetically isolated due to impassable mountains and waterfalls (Angers *et al.* 1999; Castric *et al.* 2001). Eventually, intrinsic biological characteristics can also affect connectivity, including life-history and dispersal traits, degree of ecological specialization and tolerance to environmental variations (Bringolf *et al.* 2005; Hughes *et al.* 2009; Matschiner *et al.* 2010).

The Altiplano is an intermontane endorheic basin lying between the western and eastern Cordilleras across the Central Andes of Peru, Bolivia and Chile. It spans 200,000 km^2 from 14 to 22°S, with altitudes ranging from 3700 to 4600 m (Lavenu 1992; Fornari *et al.* 2001). The region is characterized by extreme climatic conditions, with great irradiance, low temperatures, a semi-arid to arid climate, and a north-to-south gradient of annual rainfalls (~ 900 mm north to ≤ 100 mm south) (Wirrmann & Mourguiart 1995; Vila *et al.* 2007). At the present time, the central Altiplano is mainly represented by the 'TDPS' hydrological system (Lake Titicaca, Rio Desaguadero, Lake Poopo and Salar de Coipasa) in Bolivia and Peru. The southern Altiplano is represented by the Salar de Uyuni and the region of Sur Lipez in Bolivia and several smaller, high altitude salt pans ('salares'), wetlands ('bofedales') and lakes (notably in Chile) with temporary potentially subterranean connections (Vila 1975; Vila *et al.* 2013). The origin of the Altiplano dates back to the end of the Miocene (Moon 1939; Lavenu 1992). Between 1.6 Mya and 14.5 kya, this area was covered by a series of five paleolakes that underwent a succession of water-level fluctuations affected by glacial-interglacial climatic cycles (Moon 1939; Lavenu *et al.* 1984; Lavenu 1992; Fornari *et al.* 2001). Such historical fluctuations of lake levels and water connections among hydrographic systems are supposed to have played a key role in the evolution and distribution patterns of the Altiplano endemic freshwater fauna, including gastropods, amphipods and teleosteans (Lüssen *et al.* 2003; Kroll *et al.* 2012; Vila *et al.* 2013).

We present a case study on a species complex of Andean pupfishes (genus *Orestias:* Cyprinodontidae). The genus *Orestias* regroups pupfishes endemic to the high-altitude hydrological systems of the inter-Andean basin, with a main zone of distribution covering the Altiplano and the TDPS system (~ 60% of the described species are endemic to Lake Titicaca; Parenti 1984b; Vila & Pinto 1986; Vila 2006; Vila *et al.* 2011). The genus represents a unique case of freshwater teleostean diversification at high altitude. Because of its great ecological and morphological diversity, it has long been considered a species flock (Villwock 1962; Kosswig & Villwock 1964), but see Parenti (1984a). We will focus on the *agassizii* complex, which has been genetically delimited as a monophyletic group after the exclusion of the '*luteus* group' as defined by Parenti (1984b) (see Lüssen *et al.* 2003 and Esquer Garrigos *et al.* 2013). The latter includes 22 species (Parenti 1984b; Vila & Pinto 1986; Vila 2006; Vila *et al.* 2011), of which seventeen are distributed outside Lake Titicaca (including six species endemic to Chile). It is the most diverse and geographically widespread of the four complexes traditionally defined within the genus (Parenti 1984b; Lauzanne 1992; Northcote 2000; Maldonado *et al.* 2009). Taxonomic delimitations within the *agassizii* complex have been hotly debated (Parenti 1984b; Villwock 1986; Loubens 1989; Lauzanne 1992; Müller 1993; Villwock & Sienknecht 1995; Villwock & Sienknecht 1996). Three recent molecular phylogenetic studies conducted on the complex failed to recover the monophyly of any morpho-species (Lüssen *et al.* 2003; Esquer Garrigos *et al.* 2013; Vila *et al.* 2013).

Knowledge on the factors and mechanisms underlying the morphological diversification of the *agassizii* complex is very fragmentary. Some authors have suggested that Lake Titicaca has been the source for the rapid diversification of the main complexes of *Orestias* during the Pleistocene (Parenti 1981; Lauzanne 1982; Parenti 1984b; Villwock 1986; Lüssen *et al.* 2003; Maldonado *et al.* 2009), likely processing as parallel adaptive radiations (Villwock 1986; Maldonado *et al.* 2009). Nevertheless, given that the *agassizii* complex is

distributed inside and outside Lake Titicaca, its diversification and current distribution is expected to result from a combination of processes including diversification within Lake Titicaca (source) and dispersal (sinks) (Gaggiotti 1996) leading to diversification outside the lake (Parenti 1984b; Villwock 1986; Lüssen *et al.* 2003), a scenario observed in co-occurring taxa (Kroll *et al.* 2012). In this study, we use a phylogeographic / population genetic framework to test the following hypotheses related to the impact of geomorphological processes on the supposedly rapid diversification of the complex: (i) Lake Titicaca was the center of dispersion of the *agassizii* complex, (ii) genetic connectivity among populations follows the distribution of Pleistocene paleolakes on the Altiplano and episodes of water-level fluctuations among hydrographic basins, and (iii) geophysical factors such as distance, elevation and hydrographic morphology shaped patterns of isolation in the complex.

Material and Methods

Sample collection and DNA extraction

Sampling sites included 51 localities within the central and southern Altiplano from Peru, Bolivia and Chile (14 to 22° S), covering mainly the TDPS system but also various habitats such as salt pans, lakes, rivers and small wetlands. Populations were also sampled from volcanic landscapes characterized by small, closed lacustrine basins with an inflow closely controlled by ground-waters, in Bolivia (e.g. Salar de Uyuni and the region of Sur Lipez) and Chile (e.g. Salar de Ascotán and Salar de Carcote). Overall, the sampling zone represents an altitudinal gradient ranging from 3,600 to 5,000 m and seven different hydrographic basins (from A to G) not or sporadically connected to each other (e.g. during very wet seasons; Fig. 1).

A total of 822 specimens were collected from Peru and Bolivia. Given the difficulties in identifying *Orestias* species with the available identification keys, specimens were assigned

to the *agassizii* complex following the external morphological criteria given in Parenti (1984b) and Lauzanne (1982). Specimens were preserved in 99 % EtOH and deposited in the collections of Universidad Mayor de San Andres, Bolivia (UMSA). Genomic DNA was extracted using ABI Prism TM 6100 Nucleic Acid PrepStation (Applied Biosystems, Foster City, USA), following manufacturer's recommendations. For Chilean samples, 177 specimens of the *agassizii* complex were collected: *O. Ascotánensis* from Salar de Ascotán, *O. gloriae* from Salar de Carcote, *O. laucaensis* from Rio Lauca, *O. agassizii* from Salar de Huasco and Rio Isluga, and a series of specimens from Rio Lauca showing particular external morphological characteristics (*O.* sp.; MM, pers. comm.). For each specimen, a fin clip was preserved in 95% EtOH for genetic analysis. DNA extraction was done using standard salt-extraction (Aljanabi & Martinez 1997) and phenol/chloroform (Hillis *et al.* 1990) protocols.

PCR amplification and sequencing of mitochondrial genes

PCR amplification of mitochondrial genes in Bolivian and Peruvian samples targetted a 398 bp fragment of the control region (CR) using the primer pair L-Smel – H-Smel (Falk *et al.* 2003), and 1102 bp of cytochrome *b* (Cyt*b*) using the primer pair GluF – ThrR (Machordom & Doadrio 2001). PCR amplifications were performed in a 25 µl final volume with ~100 ng of template DNA, following conditions detailed in Esquer Garrigos *et al*. (2013). In the case of the Chilean samples, a fragment of 972 bp of CR was initially amplified using primers and PCR conditions as described in Morales *et al*. (2011). Sequences were then adjusted after alignment to the CR fragment size amplified for Bolivian and Peruvian samples. For Cyt*b*, a fragment of 1137 bp was amplified with the primer pair L14724 – H15915 (Xiao *et al.* 2001) under PCR conditions detailed in Vila *et al*. (2013). PCR products were sequenced at GENOSCOPE-Consortium National de Recherche en Génomique (Evry, France), EUROFINS MWG Operon (Ebersberg, Germany), and Macrogen Inc (Seoul, Republic of

South Korea), using ABI 3730XL DNA Analyzer 96-capillary sequencers (Applied Biosystems). Nucleotide sequences were edited and aligned by eye using BioEdit version 5.0.6 (Hall 1999) and deposited in GenBank (Accession numbers: xxx-xxx).

Microsatellite amplification and scoring

Nine microsatellite markers were amplified using PCR conditions as described in Esquer Garrigos *et al*. (2011). One microsatellite locus (C101) showed severe PCR amplification problems for most samples, and was not considered here. Allele scoring was done on an ABI PRISM 3100 Genetic Analyzer using internal size standard GeneScan 500 LIZ (Applied Biosystems). Allele size was determined in GENEMAPPER version 3.0 (Applied Biosystems).

Mitochondrial DNA analyses

The most likely models of molecular evolution for CR and Cyt*b* were tested from alignments pruned from indels using jModelTest version 0.1.1 (Posada 2008). Model selection was based on the Bayesian Information Criterion (Barrier *et al.* 2001), as performance analyses on simulated datasets have suggested that BIC showed a higher accuracy (Luo *et al.* 2010). The best fitting substitution model was TrN+I+G (Tamura & Nei 1993) for both genes. We used this model to build a Neighbor-Joining (NJ) tree using the concatenated data set (CR + Cyt*b*) in MEGA version 5.1 (Tamura *et al.* 2011), with default parameters. Node support was calculated using 1,000 bootstrap replicates (Felsestein 1985).

To visualize the general pattern of haplotype genealogy, we built a median-joining network using CR + Cyt*b* in Network version 4.6.1.1 (http://www.fluxus-engineering.com), with the parameter ε fixed to 0 to minimize alternative networks. Given the absence of phylogenetic and genealogical structure (see below) in our reconstructions, genetic diversity

indexes were estimated for the whole data set. We used DNAsp version 5 (Librado & Rozas 2009) to calculate number of polymorphic sites (*S*), nucleotide diversity (*Pi*), number of haplotypes (*h*) and haplotype diversity (*Hd*). Before calculations, sequences with a certain level of missing data but represented in other samples were removed from the analysis. We estimated the level of genetic differentiation among geographic populations by calculating pairwise genetic distances (Φ_{ST}) (Reynolds *et al.* 1983) in Arlequin version 3.5 (Excoffier & Lischer 2010). The significance of each Φ_{ST} values was assessed through 1,000 permutations.

We assessed past demographic events through mismatch analysis (distribution of pairwise differences) and assessed the fit of the observed distribution with the Roger's sudden expansion model (Rogers & Harpending 1992) by calculating the sum of squared deviations (SSD) between observed and expected distributions as implemented in Arlequin version 3.5, using 1,000 bootstrap replicates (Schneider & Excoffier 1999). We also calculated the Harpending raggedness index (*r*) that quantifies the smoothness of the observed pairwise distribution in DNAsp version 5. *P*-values was calculated under the coalescent with neutral infinite-sites model and large constant population sizes using 1,000 replicates (Hudson 1990). We assessed historical, demographic signatures by testing for deviation from neutrality in DNAsp version 5 using a series of statistics, following Ramirez Soriano *et al*. (2008): i) based on frequency of mutations: Fu and Li's *D*, *D** and *F** (Fu & Li 1993) and R_2 (Ramos-Onsins & Rozas 2002), and ii) based on haplotype distribution: Fu's *FS* (Fu 1997), Wall's *B* and *Q* (Wall 1999), Kelly's *ZnS* (Kelly 1997) and Roza's ZA (Rozas *et al.* 2001). *P*-values for each statistics was calculated as for *r*. The inference of demographic signatures, together with their level of intensity and relative temporality, followed the decision tables provided by Ramirez Soriano *et al.* (2008).

Analyses based on microsatellites

To calculate genetic diversity estimates from microsatellite data, we partitioned our data set following the geographic localities (n=44). Localities that were represented by low sample sizes ($n \leq 5$) were grouped with their nearest neighbor (maximal distance $\leq$ 70 km; see Table 3). Incidence of null alleles was assessed through Micro-Checker version 2.2.3 (Oosterhout *et al.* 2004). Departure from Hardy-Weinberg proportions and linkage disequilibrium were estimated in Genepop version 1.2 (Raymond & Rousset 1995). To minimize type-I errors, *P* values were adjusted for multiple tests of significance using the sequential Bonferroni correction at the 5 % nominal level (Rice 1989). We estimated genetic diversity in GenAlEx version 6.0 (Kroll *et al.* 2012) through number of alleles per locus (N_A), number of effective alleles (N_E), number of private alleles (P_A), fixation index (F_{IS}), and observed (H_O) and expected heterozygosities (H_E).

We used a Bayesian clustering method to infer population structure in STRUCTURE version 2.3.4 (Pritchard *et al.* 2000). Clustering analyses were run using the admixture ancestry model, correlated allele frequencies and no prior for sampled localities. Simulations were run five times independently, using a burn-in period of 150,000 sweeps and 200,000 MCMC iterations. The optimum number of genetic clusters (K) was selected using the maximal values of LnP (D) (the posterior probability of the data given K) as given by STRUCTURE. Analyses were performed on the Bioportal cluster (University of Oslo, Norway - http://www.bioportal.uio.no). In addition, ΔK-values based on the rate of change in log probabilities between the successive K-values were estimated following the method of Evanno *et al*. (2005) in STRUCTURE HARVESTER (Earl *et al.* 2012).

Relationships between geographic and genetic distances using mtDNA sequences and microsatellites

In order to assess correlations with genetic distances, three metrics of spatial distances were used: i) the difference in elevation between two points (altitudinal difference), ii) the

Euclidian distance (straight line distance between two points) and iii) the hydrologic distance (distance following waterways). These metrics were calculated using the packages HydroSHEDS (HS, http://hydrosheds.cr.usgs.gov/index.php: Lehner *et al*. 2008) and SRTM Water Body Data (SWBD, http://dds.cr.usgs.gov/srtm/version2_1/SWBD/), available at a 3 ARC-seconds resolution (90 m at the equator). From these data, we extracted the theoretical drainage networks following the classical hydro-geomorphology procedure (Jenson & Domingue 1988). Flow accumulation (or upstream area) was calculated from the HS flow direction layer. A constant threshold of 100 upstream cells was then applied to determine theoretical drainage networks. Thresholds were selected based upon a 'try and error' approach to ensure a realistic drainage network for the whole study area. Finally, a fusion between the extracted drainage network and the SWBD was made to maintain the surface characteristics of lentic areas and to create a lotic-lentic raster layer. In order to ensure consistency among geographical coordinates of localities and theoretical drainage networks, the former were slightly displaced (190-420 m). The elevation of each locality was then extracted individually to calculate the difference in altitude for each pair of localities. Euclidian distances were also recalculated from the coordinates (x, y) of each locality. Finally, the hydrologic distances were calculated using a cost distance function constrained to the lotic-lentic raster layer. An undifferentiated friction coefficient was used for the whole the river system. All calculations were performed in ArcGIS version 9.3 (ESRI 2011).

We assessed the correlation between the three geographic distances (altitude, Euclidian and hydrological) and genetic distances using Mantel test (Mantel 1967) in the package ade4 in R 2.2.1 (R Development Core Team 2005). F_{ST} were calculated in Arlequin 3.5 as Φ (Reynolds *et al.* 1983) for mtDNA sequences (CR + Cyt*b*) and θ (Weir & Cockerham 1984) for microsatellites. The significance of correlations was estimated with 10,000 permutations. Given that some of the localities were not interconnected (different

hydrological networks; see Fig. 1), Mantel tests including all sampled localities were performed only for altitudinal and Euclidian distances.

We tested the hypothesis of Lake Titicaca being the center of diversity from which dispersion of the *agassizii* complex occurred by setting the center of the lake as the 'starting point' from which the three geographic distance metrics and genetic diversity indices we recalculated. All the localities within Lake Titicaca were pooled and considered as a single locality. The F_{ST} values and microsatellite-based diversity indices, including effective number of alleles (N_E) and number of private alleles (P_A), were recalculated. If Lake Titicaca is the source population, we expect i) a positive correlation between genetic differentiation (measured as F_{ST}) and geographic distances, and ii) a negative correlation between genetic diversity (N_E and P_A) and geographic distances (see Eckert *et al.* 2008 and Gaubert *et al.* 2011). This approach was only applied to localities regrouping the largest hydrographic group sharing a common hydrological network ('hydrographic group A'; see Fig. 1).

Results

Phylogenetic and network reconstruction using mitochondrial DNA sequences

Our phylogenetic tree using concatenated CR + Cyt*b* sequences (*N*= 691; 1422 bp) supported the monophyly of the *agassizii* complex (Fig. 2-A). In the phylogenetic tree, no geographic structure was detected among Peruvian and Bolivian samples, with most clades showing very low bootstrap values (<75%; data not shown). The monophyly of several Chilean populations and/or species was supported, including *O. agassizii* from Salar de Huasco and Rio Collacagua, *O. ascotánensis* from Salar de Ascotán and *O. gloriae* from Salar de Carcote (Fig. 2-B and C). However, phylogenetic relationships among those clades were unresolved. The monophyly of *O. laucaensis* and *O. sp* from Rio Lauca was not supported (data not shown). The two first offshoot lineages of the complex *agassizii* had no taxonomic or

geographical consistency but were highly supported (two specimens from lake Saracocha and two specimens from Rio Lauca, respectively; both clades highlighted in red in Fig. 2-A).

The reconstructed median-joining network (n=237 haplotypes) depicted an intricate genealogical pattern, showing several haplotypes with high frequencies, and a large number of unique haplotypes at low frequencies. Twelve haplotypes were shared between Lake Titicaca and sites in the vicinity of the lake (e.g. Zapatilla, Saracocha), but also with southward sites (lakes Poopo and Uru Uru). Twenty unique haplotypes were found within Lake Titicaca (Fig. 3). Chilean populations had twenty unique haplotypes, and only one haplotype (from Rio Lauca) shared with Lake Macaya, at the western border with Bolivia (haplotype 118; Fig.3). Several unique haplotypes (i.e. not shared with Lake Titicaca) were found in lakes at high altitude (east of Lake Titicaca, in Cordillera Occidental) and in the region of Sur Lipez. The most divergent haplotypes (8 to 10 mutation steps) corresponded to lineages collected in Rio Lauca (haplotypes 2 and 3) and Lake Saracocha (Haplotypes 39 and 35) (see Fig. 3), coinciding with the two first well-supported offshoots of the *Orestias* complex on the phylogenetic reconstruction (Fig. 2).

Genetic diversity and historical demography inferred from mtDNA sequences

High levels of genetic diversity were found within *agassizii* (Table 2). Cyt*b* showed greater number of haplotypes, haplotype diversity and number of polymorphic sites, whereas nucleotide diversity was greater for CR. Very low (min=0.006) to high (max=1.0) levels of genetic differentiation were detected among sampled localities, with most of the pairwise F_{ST} values being significant ($P<0.05$) (Table S1, Supporting Information). The localities with the lowest pairwise differentiation ($F_{ST} \leq 0.09$; $P>0.05$) occurred within the TDPS system (e.g. Lake Titicaca, including nearby areas, Lake Poopo and Lake Uru Uru), among lakes in the region of 'Sur Lipez' (e.g. Lake Celeste and Lake Chipapa) and among lakes at high altitude,

northeast of Lake Titicaca (e.g. lakes in the Ulla Ulla system). The highest levels of genetic differentiation were detected among lakes and salt pans in the Chilean Altiplano (F_{ST}=0.9-1.0) and among lakes at high altitude in Bolivia (e.g. Ascotán and Catantica) and in the region of Sur Lipez (e.g. Huasco and Celeste).

Mismatch distribution did not significantly fit the Roger's sudden expansion model (SSD= 0.001, *P*= 0.569). On the other hand, *r* (0.003, *P*=0.000) and the series of statistics from Ramirez-Soriano *et al*. (2008) showed a significant deviation from neutrality, fitting a scenario of strong and relatively recent sudden expansion (Fig. 4, Table 1; see Table 3 in Ramirez-Soriano *et al*. 2008).

Genetic diversity and structure using microsatellites

The expected distribution of homozygote size classes suggested the incidence of null alleles for loci A116, C102, B103, C105, D110 and A9a, although this pattern was not consistent among all localities (Table 3). Most of microsatellite loci showed moderate to high levels of genetic diversity across the geographic zone under study. Mean number of effective alleles varied from 1.6 to 8.3, number of private alleles from 0 to 0.556, and observed and expected heterozygosity from 0.233 to 0.731 and 0.249 to 0.847, respectively (Table 3). Pairwise F_{ST} values showed low to high levels of genetic differentiation, with most of the pairwise F_{ST} values being significant (F_{ST} range = 0 to 0.5351; Table S2, Supporting Information). The patterns of genetic differentiation were similar to those found with mtDNA sequences.

Analysis with STRUCTURE suggested *K*=7 as the optimal number of genetic groups (Figs 5 and 6). Group 1 (Fig. 6-B) had the largest distribution over most of the localities within the TDPS system (except high altitude lakes) and satellite lakes of Lake Titicaca, but also outside Lake Titicaca basin in Huni (within the Siete Lagunas system Fig.1-D), Sur Lipez

(Villa Mar, Sur Lipez and Villa Alota), and north Chile (Collacagua and Huasco, and Ascotán, respectively Fig.1-B and -F). The genetic assignment of a fair proportion of specimens co-occurring with group 1 proved problematic (admixed with group 2: qi<0.80; see Fig. 6-B, C and J). Group 2 (Fig. 6-C) had the second largest distribution, from western Lake Titicaca to Lake Poopo, Lake Uru Uru, Sur Lipez region (Villa Alota) and north Chile (Rio Lauca, Huasco, Isluga, Collacagua and Ascotán). Group 3 (Fig. 6-D) included the three localities within the Hichu Khota valley (~ 4335 m; Hichu Khota, Khara Khota and Khotia), located at of altitude. Group 4 (Fig. 6-E) was restricted to six localities at high altitude (ranging from 4300 to 4800 m) north-east of Lake Titicaca (all within the Ulla Ulla system). Group 5 (Fig. 6-F) was limited to a few high altitude localities southward Bolivia in Sur Lipez (Chipapa, Celeste, Celeste Este and Chalviri). Group 6 (Fig. 6-G) included the population from Rio Empexa. Group 7 (Fig. 6-H) was restricted to high altitude lakes in the Siete Lagunas system.

Relationships between geographic and genetic distances

Mantel tests showed that geographic and genetic distances were significantly, positively correlated (P<0.001) using both mtDNA sequences and microsatellites and whatever the distance metrics (i.e. altitude and Euclidian distances for all the study zone, and altitude, Euclidian and hydrological distances within hydrographic group A; Table 4 and Fig. 7).

The distribution of pairwise geographic and genetic distances from Lake Titicaca showed a positive correlation trend (Fig. 8), which was significant only for altitude. Correlation between geographic distance and genetic diversity (N_E and P_A) showed a negative trend (Fig. 9); altitude and P_A was the only significant relationship (Table 5). The distribution of pairwise comparisons between geographic and genetic distances or genetic diversity showed a two-partition pattern roughly delimited by threshold values of 200 m (altitude) or

200 km (other distances). Within the first partition (≤200 m or km), amplitude in genetic distances and diversity was much larger, whereas in the second partition the distribution of pairwise comparisons followed a more regular positive correlation trend. Separate analysis of each partition resulted in higher and significant correlations ($P< 0.05$; data not shown).

Discussion

Lake Titicaca as the center of dispersion of the agassizii *complex?*

Lake Titicaca has been hypothesized as the center of diversification of species flocks and subsequent dispersal into southern latitudes in groups such as amphipods, microgastropods, and teleosteans (including *Orestias*) (Villwock 1962; Kosswig & Villwock 1964; Hershler & Thompson 1992; Gonzalez & Watling 2003; Väinölä *et al*. 2008). Recently, Kroll *et al*. (2012) confirmed this hypothesis for the genus *Heleobia* (micrograstropods) using a molecular phylogenetic approach. Our phylogenetic analysis of *Orestias* suggested that Lake Titicaca as the center of diversification of the genus was the most parsimonious scenario (Parenti 1984b; Villwock 1986; Lüssen *et al.* 2003). Although not directly tested in our analyses, we observed a series of evidence that support a similar 'source-sinks' pattern from Lake Titicaca in the *agassizii* complex. First, Lake Titicaca had the highest (mtDNA; data not shown) or among the highest (microsatellites; highest value in the neighboring Lake Poopo) genetic diversity across the Altiplano. It has been suggested that not only the diversity of morphotypes, but also highest levels of genetic polymorphism, are valuable indicators of centers of diversification (Rokas *et al.* 2003; Eckert *et al.* 2008). Second, the geographically most widespread genetic group included Lake Titicaca (Group 1 – microsatellites; Fig. 6), which covered surrounding satellite lakes and reached as far south as Sur Lipez region and north Chile. Considering the relative stability of the paleolakes that led to Lake Titicaca and the geomorphological conformation of central and southern Altiplano (Lavenu *et al.* 1984; Lavenu 1992; Wirrmann & Mourguiart 1995; Fornari *et al.* 2001), where Lake Titicaca is

surrounded by high altitude mountains, it is unlikely that the high genetic diversity observed in Lake Titicaca is the result of a high level of admixture among different source populations having dispersed into the lake (reservoir function). As such, we consider that the divergent mtDNA haplotypes found in Lake Saracocha and Rio Lauca (Fig. 2-A) are climate relicts from a past metapopulation (Hampe & Jump 2011). Third, using microsatellites and Lake Titicaca as a starting point, we found a positive correlation between geographic distances (altitude, Euclidian, hydrological) and genetic distances, and a negative correlation with genetic diversity. Such trends fit with the expected pattern of source-sink dynamics from Lake Titicaca, i.e. dispersal from the lake into satellite and southern localities (Eckert *et al.* 2008; Gaubert *et al.* 2011) . Eventually, the absence of mtDNA structure within and outside Lake Titicaca, together with the occurrence of a fair proportion of admixed genotypes, argue for recent interconnections within the TDPS system (notably between genetic groups 1 and 2; Fig. 6) and among the TDPS and other basins of the Altiplano.

Assigning Lake Titicaca as the center of origin of the *agassizii* complex may raise two issues. First, the morphological diversification of the complex within the lake appears weak (5 described species) compared to the 17 species found outside the lake (Parenti 1984b; Vila & Pinto 1986; Vila 2006; Vila *et al.* 2011). Such pattern contrasts with that found in other *Orestias* complexes, for which most or all species are endemic to Lake Titicaca. In this case, it is conceivable that either (i) taxonomic malpractices based on difficult material (see Esquer Garrigos *et al.* 2013) or (ii) severe extinction events in Lake Titicaca, have forged the morpho-species distribution pattern currently observed within the *agassizii* complex. Second, the *agassizii* complex is the sole complex genuinely distributed outside Lake Titicaca; although *O. luteus* can be found in satellite lakes of Lake Titicaca and in Lake Uru Uru (Parenti 1984b; Esquer Garrigos *et al.* 2011). It is possible that some intrinsic biological attributes of the *agassizii* complex allowed the latter to be more 'ecologically successful'.

Indeed, the complex spreads as far as 1285 km south (Celeste Este; hydrographic distance) from Lake Titicaca across a wide spectrum of habitats with large variations in salinity (average= 1 $g.l^{-1}$) and mean monthly temperature (5-10 °C) (Vila *et al.* 2007). The diversification and subsequent dispersal of other species flocks across central and southern Altiplano (e.g. *Helobia*, Kroll *et al*. 2012) have been hypothesized to be linked to a wide spectrum of salinity tolerance, which permitted to ancestral populations to tolerate severe changes in salinity during lake level fluctuations. The sister clade of the genus *Orestias* contains species living in estuarine or highly mineralized waters (Villwock 1986; Parker & Kornfield 1995; Berra 2001; Nelson 2006). It is possible that this ability has persisted or re-occurred in the *agassizii* complex, thus allowing for greater dispersal capacities across central and southern Altiplano eventually leading to allopatric diversification. Moreover, representatives of the *agassizii* complex are often considered as having a more generalist diet than species from other complexes (although some degree of trophic niche overlap exists; Lauzane 1992; Loubens 1989; Northcote 2000; Maldonado *et al*. 2009). Since generalist species have a tendency to be better dispersers than specialists, we hypothesize that this life-history trait could also promote the spread and diversification of the *agassizii* complex through central and southern Altiplano. However, it will be important to better characterize the ecology and dispersal abilities of members of the complex to assess whether or not those factors significantly influenced their current ecological success.

Impact of paleolake fluctuations and geomorphological characteristics on the genetic structure of the agassizii *complex*

Our results based on the analysis of microsatellites (STRUCTURE) tend to illustrate a major impact of paleolake fluctuations rather than current basin delimitations on population connectivity within the *agassizii* complex. One of our most striking results is the almost

optimal superimposition of the maximum extent of Paleolake Ballivian (Fig. 1 in Kroll *et al*. 2012; Fig. 6-A in this study) with the geographic delimitation of the genetic group 1, including the TDPS system except high altitude lakes, northern Sur Lipez and Chile (Collacagua, Ascotán and Huasco). The extension of the Paleolake Ballivian occurred approx. between 0.6 and 0.5 Mya (Lavenu *et al.* 1984; Lavenu 1992; Wirrmann & Mourguiart 1995), and has been associated to the dispersal and diversification of the *agassizii* complex (Lüssen *et al.* 2003) and the gastropod species flock *Heleobia* (Kroll *et al.* 2012). Importantly, this extension could have persisted recently through Paleolake Tauca (ca. 18-14.5 kya; Lavenu *et al*. 1984 and Lavenu 1992). Moreover, we found across the group 1's range a fair level of admixed individuals with the genetic group 2 (distributed from western Lake Titicaca to Lake Poopo, Lake Uru Uru, western border with Chile, and Isluga and Lauca in Chile). This suggests that the division of paleolakes into two water bodies by the Ulloma-Callapa sill in the Rio Desaguadero (Lavenu *et al.* 1984; Lavenu 1992), probably after the Ballivian episode, may have allowed a north-south differentiation within the TDPS area, and that subsequent interconnections —via Rio Desaguadero, or during high rainfalls between Lake Poopo basin and Salar de Coipasa, and Salar de Uyuni and northern Sur Lipez (Roche *et al.* 1992; Wirrmann & Mourguiart 1995)— yielded the admixed pattern currently observed between the two genetic groups. Another possibility is that the genetic group 2 dispersed from the western Altiplano (where non-admixed group 2 populations occur) to the north and east during the maximum extension of Paleolake Ballivian, although a reverse dispersal from populations differentiated south of TDPS to western central Altiplano is also possible (notably through flooding waters from Lake Poopo to Salar de Coipasa and Salar de Uyuni; Blodgett *et al*. 1997). The apparent contradicting pattern of 'pure' genetic group 2 populations at P. Opoqueri, which is located within the presumable delimitation of Lake Ballivian, is explained

by the fact that it is an artificial lake, recently created during a road network construction work by diverting the watercourse of Rio Lauca (probably in 2009; CI, pers. comm.).

Bayesian clustering analysis suggested that basin delimitation within the Altiplano had little effect on the genetic structure of the *agassizii* complex (Fig. 8). We posit that genetic groups 3 to 7 dispersed from the Paleolake Ballivian metapopulation at the maximum water level reached by the paleolake. Rio Empexa populations (genetic group 6) were probably isolated after decrease in water level resulting in the end of the connexion with Paleolake Tauca (Lavenu *et al*. 1984 and Lavenu 1992). Sur Lipez populations (genetic group 5) may have originated from dispersal at the southern maximal limit of paleolakes Ballivian and Tauca, followed by habitat fragmentation and separation from northward areas including Salar de Coipasa and Salar de Uyuni. Indeed, during the Quaternary, Sur Lipez experienced intense volcanic activities with very dry climate conditions that affected hydrological systems and contributed to the formation of small wetlands ('bofedales') isolated by unsuitable habitats. Siete Lagunas (genetic group 7), Ulla Ulla (genetic group 4) and Hichu Khota valley (genetic group 3) hold populations isolated from the TDPS system by a steep altitudinal gradient (ca. 600 to 1000 m). For this reason, it remains difficult to explain the dispersal of a group with such poor swimming abilities and restricted to lentic ecosystems as the *agassizii* complex.

Correlation between genetic and geographic distances and possible scenario of diversification of the agassizii *complex in the Altiplano*

Our analyses showed that geographic and genetic distances were significantly, positively correlated using both mtDNA sequences and microsatellites, and whatever the distance metrics used (Mantel test; Table 4 and Fig. 9). This suggests a strong effect of isolation-by-distance (Wright 1943) across the Altiplano. Moreover, pairwise distributions of geographic

and genetic distances or diversity from Lake Titicaca showed a two-partition pattern delimited by a 200 km (distances) threshold value (altitude dealt with below). Apparently, isolation-by-distance was a major force shaping the genetic structure of the complex around Lake Titicaca basin (which roughly corresponds to a radius of 200 km from the "center" of the lake; first partition), despite regular flooding of the basin that could have promoted a certain genetic 'homogeneity' of the area. Within the second partition (≥200 km), where flooding from Lake Titicaca was sporadic (Lavenu *et al.* 1984), distance from the source also had a significant influence (positive correlation), making possible a scenario of diversification "by force of distance" outside the lake (Irwin *et al.* 2005). Euclidian and hydrological distances are two extrinsic factors known to affect connectivity (dispersal) and genetic structure in freshwater ecosystems (Koblmüller *et al.* 2007; Hughes *et al.* 2009; Matschiner *et al.* 2010). Unfortunately, in the case of the *agassizii* complex, the dynamics and routes of connections among populations and basins are poorly known, even at the present time (refs). Possibly, in Chile and Sur Lipez (southern Altiplano), contemporary connections are subterranean (Vila 1975; Vila *et al.* 2013).

Remarkably, elevation had the most significant effect on genetic distance and diversity distribution, independent of the Euclidian or hydrographic distances separating populations (Table 5 and Figs. 8 and 9). This observation was also supported by the Bayesian inference of genetic structure, which showed that lake systems at high altitudes (ca. 4100-4800 m) were generally well differentiated, including Siete Lagunas (south-west Lake Titicaca, outside Lake Titicaca basin), Hichu Khota and Ulla Ulla (north-east Lake Titicaca, within Lake Titicaca basin). The role of elevation was also underlined through the highest F_{ST} values found between pairs including high altitude systems. The variation of elevation across a given range is known to have a significant structuring effect on freshwater populations because of the occurrence of impassable barriers such as mountains, waterfalls and 'steep slopes' (Angers *et*

al. 1999; Castric *et al.* 2001). In such cases, high levels of genetic differentiation and microscale endemism can be explained by a greater effect of genetic drift and inbreeding on isolated populations with small effectives (Frankham 1998), but also by divergent local adaptation in new, restricted habitats (Hughes *et al.* 2009; Cristescu *et al.* 2012). Our results suggest that there is a high potential for long distance dispersal —whether active or not— and colonisation of new habitats in the *agassizii* complex. This poses the question of how members of the complex managed to reach localities sometimes reaching 1000 m above Lake Titicaca. Indeed, Andean pupfishes are known to be poor swimmers and to prefer stagnant water bodies. Although underground hydrosystems may constitute conditions that promoted the dispersal of pupfishes across central and southern Altiplano (see above), their capacity of dispersing through river flows and underground rivers, and to resist low temperatures remains to be characterized.

The genetic structure and dispersal pathways that we have started to explore in the *agassizii* complex make us envisage a preliminary scenario of diversification across central and southern Altiplano. The demographic analysis based on mtDNA sequences detected a strong and relatively recent sudden expansion event within the complex that fits a scenario of rapid diversification and/or geographic dispersal (unimodal mismatch distribution; Ray *et al.* 2003) during the Ballivian-Tauca period (0.6 Mya-14.5 kya). In the *agassizii* complex, the source from which diversification and dispersals occurred would have been the paleolakes at the origin of Lake Titicaca, with Paleolake Ballivian playing a significant role during its maximum extent in allowing members of the complex to disperse into southern localities (Sur Lipez, western Altiplano and Chile). Later, isolation-by-distance effect (Irwin *et al.* 2005) coupled to subsequent isolation following drier periods would have made those populations genetically divergent. As such, the available evidence does not refute the hypothesis of a scenario combining ecological speciation within Lake Titicaca (diversification with inter-

specific competition) and allopatric speciation outside the lake (diversification without inter-specific competition) (Parenti 1984b; Villwock 1986; Lüssen *et al.* 2003; Vila *et al.* 2013). However, the genetic structure observed (microsatellites) argues for recent long-distance inter-connections among basins, notably between the TDPS system, Sur Lipez and Empexa (Bolivia), suggesting that the dynamics of dispersal and diversification is still active within the complex.

Conclusion

Our preliminary study suggested a rapid diversification of the *agassizii* complex across central and southern Altiplano shaped by paleolake dynamics and episodes of water-level fluctuations among hydrographic basins, involving the long-distance dispersal of currently isolated and divergent populations at high altitude. Nevertheless, the addition of remote populations from Peru, not included in our study, may render our scenario more complex.

In terms of conservation, our results based on microsatellite loci determined local evolutionary significant units (ESUs; Moritz 1994) restricted to fragile habitats that might constitute important targets for future conservation plans given the extinction-proneness of such populations (Moritz 1994; Frankham 1998): (i) the Siete Lagunas system, with four of the seven lakes (three of which included in our study) having recently dried out; (ii) the Hichu Khota valley, (iii) the Ulla Ulla system, the only ESU benefiting from a protected range (Ulla Ulla National Reserve); (iv) Rio Empexa; and (v) the southern Sur Lipez, characterized by a volcanic and arid landscape with isolated patches of fragile suitable habitats (small wetlands 'bofedales' isolated by unsuitable salt pans).

The Chilean species of the *agassizii* complex have been considered by some authors as representing ontogenetic variants of the species *O. agassizii* with no reproductive isolation (Villwock & Sienknecht 1995; Villwock & Sienknecht 1996) and without clear phylogenetic

definition (Lüssen *et al.* 2003; Vila *et al.* 2013). Our mtDNA-based analyses showed a certain degree of genetic clustering for Chilean morpho-species (but monophyly was generally weakly supported). On the other hand, microsatellites-based analyses failed to recover distinct genetic populations for any of the Chilean morpho-species (although in a few cases, this may be related to a certain level of missing data in our genotypes). Thus, additional investigations will be necessary to assess the biological status of the Chilean taxa with complementary data (Vila *et al.* 2010).

Human-induced translocations may act as a significant disruptor of dispersal processes and distribution patterns (Leprieur *et al.* 2008). Artificial translocations of representatives of the *agassizii* complex may have occurred and still occur across the Altiplano (e.g. Laguna Huni near the Siete Lagunas System). It thus appears urgent to control programs of translocations to ensure the genetic integrity of the ESUs that we have identified across the Altiplano.

References

Aljanabi, S.M. & Martinez, I. (1997) Universal and rapid salt–extraction og high quality genomic DNA for PCR–based techniques. *Nucleic Acids Research*, 25, 4692–4693.

Angers, B., Magnan, P., Plante, M. & Bernatchez, L. (1999) Canonical correspondence analysis for estimating spatialand environmental effects on microsatellite gene diversityin brook charr (*Salvelinus fontinalis*). *Molecular Ecology*, 8, 1043–1053.

Barrier, M., Robichaux, R.H. & Purugganan, M.D. (2001) Accelerated regulatory gene evolution in an adaptive radiation. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 10208–10213.

Berra, T.M. (2001) *Freshwater fish distribution*. Academic Press, San Diegp, 604 pp.

Bringolf, R.B., Kwak, T.J., Cope, W.G. & Larimore, M.S. (2005) Salinity tolerance of flathead catfish: implications for dispesal of introduced populations. *Transactions of the American Fisheries Society*, 134, 927–936.

Blodgett, T.A., Lenter, J.D., Isacks, B.L. (1997) Constraints on the origin of paleolakes expansion in the central Andes, *Earth Interactions*, 1, 1–28.

Castric, V., Bonney, F. & Bernatchez, L. (2001) Landscape structure and hierarchical genetic diversity in the brook charr, *Salvelinus fontinalis*. *Evolution*, 55, 1016–1028.

Costello, A.B., Down, T.E., Pollard, S.M., Pacas, C.J. & Taylor, E.B. (2003) The influence of history and contemporary stream hydrology on the volution of genetic diversity within species: an examination of microsatellite DNA variation in bull trout, Salvelinus confluentus (Pisces: Salmonidae). *Evolution*, 57, 328–344.

Coyne, J.A. & Orr, H.A. (2004) *Speciation*. Sunderland, Massachutes, 545 pp.

Cristescu, M.E., Constantin, A., Bock, D.G., Caceres, C.E. & Crease, T.J. (2012) Speciation with gene flow and the genetics of habitat transitions. *Molecular Ecology*, 21, 1411–1422.

Earl, Dent, A., vonHoldt & Bridgett, M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359–361.

Eckert, C.G., Samis, K.E. & Lougheed, S.C. (2008) Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology*, 17, 1170–1188.

Esquer Garrigos, Y., Hugueny, B., Koerner, K., Ibañez, C., Bonillo, C., Pruvost, P., Causse, R., Cruaud, C. & Gaubert, P. (2013) Non–invasive ancient DNA protocol for fluid–preserved specimens and phylogenetic systematics of the genus *Orestias* (Teleostei: Cyprinodontidae). *Zootaxa*, 3640, 373–394.

Esquer Garrigos, Y., Lambourdiere, J., Ibañez, C. & Gaubert, P. (2011) Characterization of ten polymorphic microsatellite loci in the Andean pupfish *Orestias agassizii*, with cross–amplification in the sympatric *O. luteus*. *Conservation Genetics Resources*, 3, 17–19.

ESRI. (2011) ArcGIS Desktop: Release 10. Environmental Systems Reasearch Institute, Redlands, CA.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, 14, 2611–2620.

Excoffier, L. & Lischer, H.E.L. (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564–567.

Falk, T.M., Teugels, G.G., Abban, E.K., Villwock, W. & Renwrantz, L. (2003) Phylogeographic patterns in populations of the black–chinned tilapia complex (Teleostei, Cichlidae) from coastal areas in West Africa: support for the refuge zone theory. *Molecular Phylogenetics and Evolution*, 27, 81–92.

Felsestein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39, 783–791.

Fornari, M., Risacher, F. & Féraud, G. (2001) Dating of paleolakes in the central Altiplano of Bolivia. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 172, 269–282.

Frankham, R. (1998) Inbreeding and Extinction: Island Populations. *Conservation Biology*, 12, 665–675.

Fu, Y.X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147, 915–925.

Fu, Y.X. & Li, W.H. (1993) Statistical tests of neutrality of mutations. *Genetics*, 133, 693–709.

Gaggiotti, O.E. (1996) Population Genetic Models of Source–Sink Metapopulations. *Theoretical Population Biology*, 50, 178–208.

Garant, D., Forde, S.E. & Hendry, A.P. (2007) The murltifarious effects of dispersal and gene flow on contemporary adaptation. *Functional Ecology*, 21, 434–443.

Gaubert, P., Machordom, A., Morales, A., López–Bao, J.V., Veron, G., Amin, M., Barros, T., Basuony, M., Djagoun, C.A.M.S., San, E.D.L., Fonseca, C., Geffen, E., Ozkurt, S.O., Cruaud, C., Couloux, A. & Palomares, F. (2011) Comparative phylogeography of two African carnivorans presumably introduced into Europe: disentangling natural versus human–mediated dispersal across the Strait of Gibraltar. *Journal of Biogeography*, 38, 2–17.

Gonzalez, E. R., and Watling, L. (2003) A new species of *Hyalella* from the Patagonia, Chile with re-description of *H. simplex* Schellenberg, 1943 (Crustacea: Amphipoda). *Journal of Natural Hisory*, 37, 2077–2094.

Hall, T. (1999) BioEdit: a user–friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95–98.

Hampe, A. & Jump, A.S. (2011) Climatic Relicts: Past, Present, Future. *Annual Review of Ecology, Evolution and Systematics*, 42, 313–333.

Hershler, R. and Thompson, F., G. (1992) A review of the aquatic gastropod subfamily Cocliopinae (Prosobranchia: Hydrobiidae. *Malacol Review*, 5, 1–140.

Hillis, D.M., Larson, A., Davis, S.K. & Zimmer, E.A. (1990) Nucleic acids III: sequencing. *In:* Hillis, D.M. & Mortiz, C. (Eds.) *Molecular Systematics*. Sinauer Associates, Sunderland, MA, pp. 318–370.

Hudson, R.R. (1990) Gene genealogies and the coalescent process *In:* Antonovics, J. & Futuyama, D. (Eds.) *Oxford Surveys in Evolutionary Biology*. Oxford University Press, Oxford, pp. 1–44.

Hughes, J.M., Schmidt, D.J. & Finn, D.S. (2009) Genes in Streams: Using DNA to Understand the Movement of Freshwater Fauna and Their Riverine Habitat. *BioScience*, 59, 573–583.

Irwin, D.E., Bensch, S., Irwin, J.H. & Price, T.D. (2005) Speciation by Distance in a Ring Species. *Science*, 307, 414–416.

Kelly, J.K. (1997) A test of neutrality based on interlocus associations *Genetics*, 146, 1197–1206.

Koblmüller, S., Sefc, K., Duftner, N., Warum, M. & Sturmbauer, C. (2007) Genetic population structure as indirect measure of dispersal ability in a Lake Tanganyika cichlid. *Genetica*, 130, 121–131.

Kocher, T.D. (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics*, 5, 288–298.

Kosswig, C. & Villwock, W. (1964) Das problem der intralakustrischen speziation im Titicaca und Lanaosee. *Verhandlungen der Deutschen Zoologischen Gesellschaft*, 28, 95–102.

Kroll, O., Hershler, R., Albrecht, C., Terrazas, E.M., Apaza, R., Fuentealba, C., Wolff, C. & Wilke, T. (2012) The endemic gastropod fauna of Lake Titicaca: correlation between molecular evolution and hydrographic history. *Ecology and Evolution*, 2, 1517–1530.

Lauzanne, L. (1982) Les *Orestias* (Pisces, Cyprinodontidae) du Petit lac Titicaca. *Revue d'Hydrobiologie Tropicale*, 15, 39–70.

Lauzanne, L. (1992) Lake Titicaca. A synthesis of limnological knowledge. Fish fauna. Native species. The Orestias. *Monographiae Biologicae*, 68, 405–419.

Lavenu, A. (1992) Formation and geological evolution. *In:* Dejoux, C. & Iltis, A. (Eds.) *Lake Titicaca: A synthesis of limnological knowledge*. Kluwer Academic Publishers, London, pp. 3–15.

Lavenu, A., Fornari, M. & Sebrier, M. (1984) Existence de deux nouveaux épisodes lacustres quaternaires dans l'Altiplano péruvo–bolivien. *Cahiers ORSTROM série Géologie*, 14, 103–114.

Leprieur, F., Beauchard, O., Blanchet, S., Oberdorff, T. & Brosse, S. (2008) Fish invasions in the world's river systems: when natural processes are blurred by human activities. *PLoS Biology*, 6, 0404–0410.

Librado, P. & Rozas, J. (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451–1452.

Loubens, G. (1989) Observations sur les poissons de la partie Bolivienne du lac Titicaca. 4. *Orestias* spp., *Salmo gairdneri* et problemes d'amenagement. *Revue d'Hydrobiologie Tropicale*, 22, 157–177.

Luo, A., Qiao, H., Zhang, Y., Shi, W., Ho, S., Xu, W., Zhang, A. & Zhu, C. (2010) Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. *BMC Evolutionary Biology*, 10, 242.

Lüssen, A., Falk, T. & Villwock, W. (2003) Phylogenetic patterns in populations of Chilean species of the genus *Orestias* (Teleostei: Cyprinodontidae): Results of mitochondrial DNA analysis. *Mol Phylogenet Evol*, 29, 151–160.

Machordom, A. & Doadrio, I. (2001) Evidence of a Cenozoic Betic–Kabilian Connection Based on Freshwater Fish Phylogeography (Luciobarbus, Cyprinidae). *Molecular Phylogenetics and Evolution*, 18, 252–263.

Maldonado, E., Hubert, N., Sagnes, P. & De Mérona, B. (2009) Morphology–diet relationships in four killifishes (Teleostei, Cyprinodontidae, *Orestias*) from Lake Titicaca. *J Fish Biol*, 74, 502–520.

Matschiner, M., Hanel, R. & Salzburger, W. (2010) Phylogeography and speciation processes in marine fishes and fishes from large freshwater lakes. *In:* Rutgers, D.S. (Ed.) *Phylogeography*. Nova Science Publishers, Inc1–29.

McGlashan, D.J. & Hugues, J.M. (2002) Extensive genetic divergence among populations of the Australian freshwater fish, *Pseudomugil signifer* (Pseudomugilidae), at different hierarchical scales. *Marine and Freshwater Research*, 53, 1–11.

Moon, H.P. (1939) III. The geology and physiography of the Altiplano of Peru and Bolivia. *Transactions of the Linnean Society of London: 3rd Series*, 1, 27–43.

Morales, P., Vila, I. & Poulin, E. (2011) Genetic structure in remnant populations of an endangered cyprinodontid fish, *Orestias ascotanensis*, endemic to the Ascotan salt pan of the Altiplano. *Conservation Genetics*, 12, 1639–1643.

Moritz, C. (1994) Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology & Evolution*, 9, 373–375.

Müller, R. (1993) Critical remarks on the revision of the genus *Orestias* (Pisces Cyprinodontidae) by Parenti (1984). *Zoologische Jahrbuecher Abteilung fuer Anatomie und Ontogenie der Tiere*, 123, 31–58.

Nelson, J.S. (2006) *Fishes of the World*. John Willey & Sons, New Jersey, 642 pp.

Northcote, T.G. (2000) Ecological interactions among an Orestiid (Pisces: Cyprinodontidae) species flock in the littoral zone of lake Titicaca. *Advances in ecological research*, 31, 399–420.

Oosterhout, C., Hutchinson, W., Wills, D. & Shipley, P. (2004) MICRO–CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4, 535–538.

Parenti, L.R. (1981) A phylogenetic and biogeographic analysis of cyprinodontiform fishes (Teleostei, Atherinomorpha) *Bulletin of the American Museum of Natural History*, 168, 335–557.

Parenti, L.R. (1984a) Biogeography of the Andean killifish genus *Orestias* with comments on the species flock concept. *In:* Echelle, A.A. & Kornfield, I. (Eds.) *Evolution of fish species flocks.*85–92.

Parenti, L.R. (1984b) A taxonomic revision of the Andean killifish genus *Orestias* (Cyprinodontiformes, Cyprinodontidae). *Bullutin of the American Museum of Natural History*, 178, 107–214.

Parker, A. & Kornfield, I. (1995) Molecular perspective on evolution and zoogeography of cyprinodontid killifishes (Teleostei; Atherinomorpha). *Copeia*, 1995, 8–21.

Posada, D. (2008) jModelTest: phylogenetic model averaging *Molecular Biology and Evolution*, 25, 1253–1256.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, 155, 945–959.

R Development Core Team. (2005) R: A language and environment for statistical computing, reference index version 2.2.1. Vol. 2013. R Foundation for Statistical Computing, Viena, Austria

Ramirez–Soriano, A., Ramos–Onsins, S.E., Rozas, J., Calafell, F. & Navarro, A. (2008) Statistical Power Analysis of Neutrality Tests Under Demographic Expansions, Contractions and Bottlenecks With Recombination. *Genetics*, 179, 555–567.

Ramos–Onsins, S., E & Rozas, J. (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, 19, 2092–2100.

Ray, N., Currat, M. & Excoffier, L. (2003) Intra–Deme Molecular Diversity in Spatially Expanding Populations. *Molecular Biology and Evolution*, 20, 76–86.

Raymond, M. & Rousset, F. (1995) GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86, 248–249.

Reynolds, J., Weir, B.S. & Cockerham, C.C. (1983) Estimation for the coancestry coefficient: basis for a short–term genetic distance. *Genetics*, 105, 767–779.

Rice, W.R. (1989) Analyzing tables of statistical tests. *Evolution*, 43, 223–225.

Roche, M.A., Bourges, J., Cortes, J. & Mattos, R. (1992) Climatology and hydrology of the Lake Titicaca basin. *In:* Dejoux, C. & Iltis, A. (Eds.) *Lake Titicaca: A syntesis of Limnological knowledge*. Kluwer Academic Publishers, Kluwer, Dordrecht, pp. 63–88.

Rogers, A.R. & Harpending, H. (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, 9, 552–569.

Rokas, A., Atkinson, R.J., Webster, L., Csóka, G. & Stone, G.N. (2003) Out of Anatolia: longitudinal gradients in genetic diversity support an eastern origin for a circum–Mediterranean oak gallwasp *Andricus quercustozae*. *Molecular Ecology*, 12, 2153–2174.

Rozas, J., Gullaud, M., Blandin, G. & Aguade, M. (2001) DNA variation at the rp49 gene region of *Drosophila simulans* evolutionary inferences from an unusual haplotype structure. *Genetics*, 158, 1147–1155.

Rundle, H.D. & Nosil, P. (2005) Ecological speciation. *Ecology Letters*, 8, 336–352.

Schneider, S. & Excoffier, L. (1999) Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*, 152, 1079–1089.

Seehausen, O. (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, 19, 198–207.

Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, 10, 512–526.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutinary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, 28, 2731–2739.

Unmack, P.J. (2001) Biogeography of Australian freshwater fishes. *Journal of Biogeography*, 28, 1053–1089.

Vainölä, R., Witt, J. D. S., Grabowski, M., Bradbury, J. H., Jazdzewski, K., and Sket, B. (2008) Global diversity of amphipods (Amphipoda; Crustacea) in freshwater. *Hydrobiologia*, 595, 241–255.

Vila, G. (1975) Geología de los depositos salinos Andinos, Provincia de Antofagasta, Chile. *Revista Geologica de Chile*, 2, 41–55.

Vila, I. (2006) A new species of killifish in the genus *Orestias* (Teleostei: Cyprinodontidae) from the Southern High Andes, Chile. *Copeia*, 2006, 472–477.

Vila, I., Morales, P., Scott, S., Poulin, E., Véliz, D., Harrod, C. & Méndez, M.A. (2013) Phylogenetic and phylogeographic analysis of the genus *Orestias* (Teleostei: Cyprinodontidae) in the southern Chilean Altiplano: the relevance of ancient and recent divergence processes in speciation. *Journal of Fish Biology*, 82, 927–943.

Vila, I., Pardo, R. & Scott, S. (2007) Freshwater fishes of the Altiplano. *Aquatic Ecosystem Health & Management*, 10, 201–211.

Vila, I. & Pinto, M. (1986) A new species of Killifish (Pisces, Cyprinodontidae) from the Chilean Altiplano. *Revue d'Hydrobiologie Tropicale*, 19, 233–239.

Vila, I., Scott, S., Lam, N., Iturra, P. & Méndez, M.A. (2010) Karyological and morphological analysis of divergence among species of the killifish genus *Orestias* (Teleostei: Cyprinodontidae) from the southern Altiplano. *In:* Nelson, J.S., Schultze, H.P. & Wilson, M.V.H. (Eds.) *Origin and Phylogenetic Interrelationships of Teleosts*, Munich, Verlag, pp. 471–480.

Vila, I., Scott, S., Mendez, M.A., Valenzuela, F., Iturra, P. & Poulin, E. (2011) *Orestias gloriae*, a new species of cyprinodontid fish from saltpan spring of the southern high Andes (Teleostei: Cyprinodontidae). *Ichthyological Exploration of Freshwaters*, 22, 345–353.

Villwock, W. (1962) Die Gattung *Orestias* (Pisces, Microcyprini) und die Frage der intralakustrischen Speziation im Titicaca–Seengebiet. *Verhandlungen der Deutschen Zoologischen Gesellschaft Wien*, No. 54, 610–624.

Villwock, W. (1986) Speciation and adaptive radiation in the Andean *Orestias* fishes. *In:* Vuilleumier, F. & Monasterio, M. (Eds.) *High altitude tropical biogeography*. Oxford University Press, Oxford, pp. 387–403.

Villwock, W. (1993) Lake Titicaca region on the Altiplano of Peru and Bolivia and the consequences of introduced fish for wild species and their environment. *Die Titicaca–see–region auf del Altiplano von Peru und Bolivien und die folgen eingefuhrter fische fur wild–arten und ihren lebensraum*, 80, 1–8.

Villwock, W. & Sienknecht, U. (1995) Intraspezifische variabilität im genus *Orestias* Valenciennes, 1839 (Teleostei: Cyprinodontidae) und zum problem der artidentität. *Mitteilungen aus dem Hamburgischen Zoologischen Museum und Institut*, 92, 381–398.

Villwock, W. & Sienknecht, U. (1996) Contribution to the knowledge and history of Chilean fishes. The cyprinodontids of the genus *Orestias* Val. 1839 (Teleostei: Cyprinodontidae) of the Chilean Altiplano. *Medio Ambiente*, 13, 119–126.

Wall, J.D. (1999) Recombination and the power of statistical tests of neutrality. *Genetic Research*, 74, 65–79.

Weir, B.S. & Cockerham, C.C. (1984) Estimating F–statistics for the analysis of population structure, *Evolution*, 38, 1358–1370.

Wirrmann, D. & Mourguiart, P. (1995) Late Quaternary spatio–temporal limnological variations in the Altiplano of Bolivia and Peru. *Quaternary Research*, 43, 344–354.

Wong, B.B.M., Keogh, J.S. & McGlashan, D.J. (2004) Current and historical patterns of drainage connectivity in eastern Australia inferred from population genetic structuring in a widespread freshwater fish *Pseudomugil signifer* (Pseudomugilidae). *Molecular Ecology*, 13, 391–401.

Wright, S. (1943) Isolation by distance. *Genetics*, 28, 114–138.

Xiao, W.H., Zhang, Y.P. & Liu, H.Z. (2001) Molecular systematics of *Xenocyprinae* (Teleostei: Cyprinidae): taxonomy, biogeography, and coevolution of a special group restricted in East Asia. *Molecular Phylogenetics and Evolution*, 18, 163–173.

Figures.

Figure 1. Sampling localities. Localities interconnected through a common hydrographic network are shown in same color ('hydrographic groups'). Identifier (ID) for each locality or group of localities is given (see Table 3). Basins are delimited by the thick black contour.

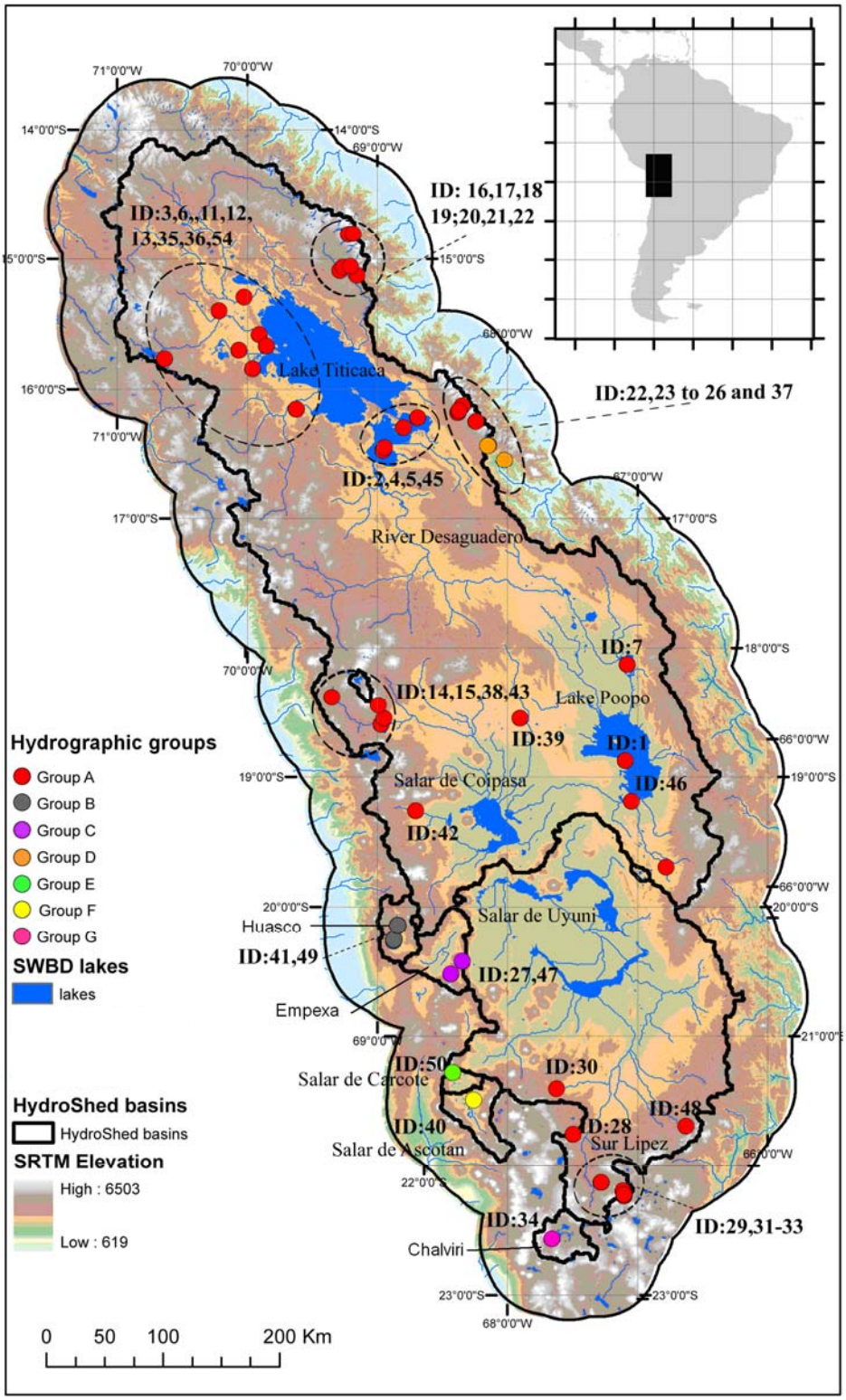


Figure 2. Neighbor-joining tree of concatenated mtDNA sequences (CR + Cyt*b*), showing A) the general pattern of the reconstructed tree (ingroup: *agassizii* complex); the two colored branches correspond to haplotypes from Lake Saracocha and Rio Lauca, and B) the recovered clades from Chile (B and C). Values at branch nodes are bootstrap estimates ≥80. Scale bars below trees indicate proportion of sequence divergence.

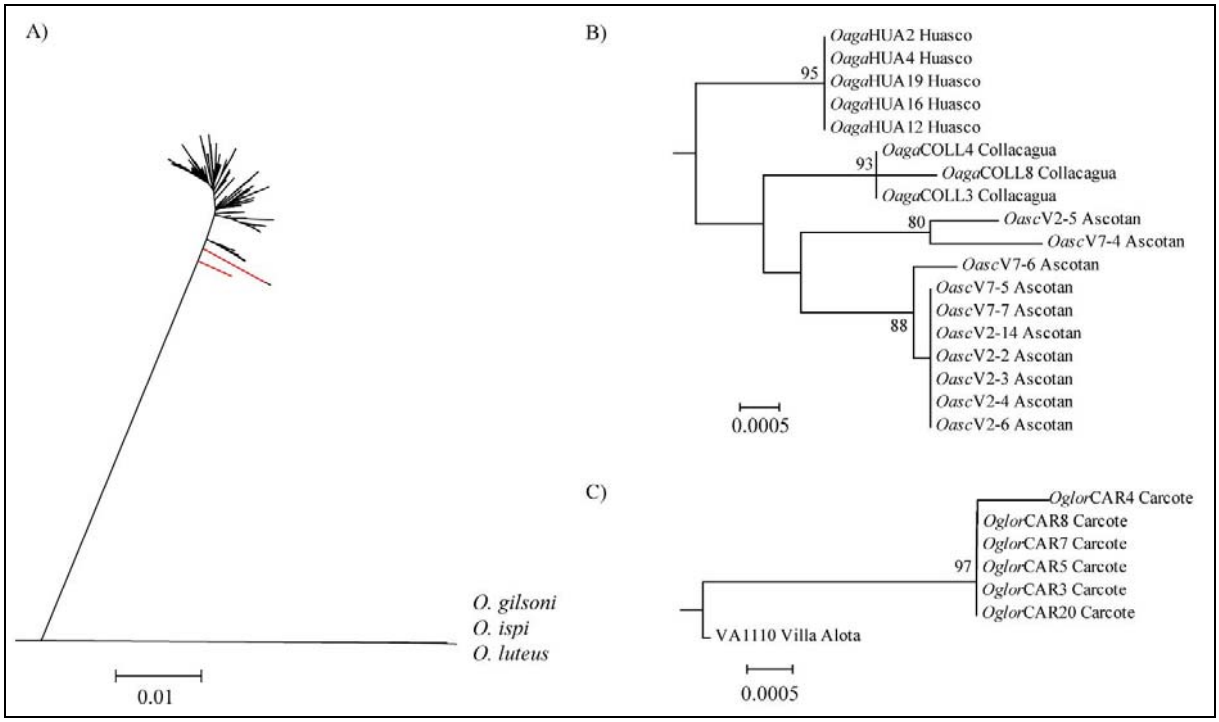


Figure 3. Median joining network of concatenated mtDNA sequences (CR + Cyt*b*) for the *agassizii* complex, showing the general structure of the network without (A) and with (B) the frequency of haplotypes. In B, circle size is proportional to the number of individuals sharing the same haplotype. Scale bar indicates one mutational step. Highlighted haplotypes, in green: haplotypes found in and outside Lake Titicaca; blue: unique haplotypes found in Lake Titicaca, pink: unique Chilean haplotypes — H_118 is shared between Isluga (Chile) and Macaya (Bolivia). Yellow haplotypes are the other haplotypes found outside Lake Titicaca.

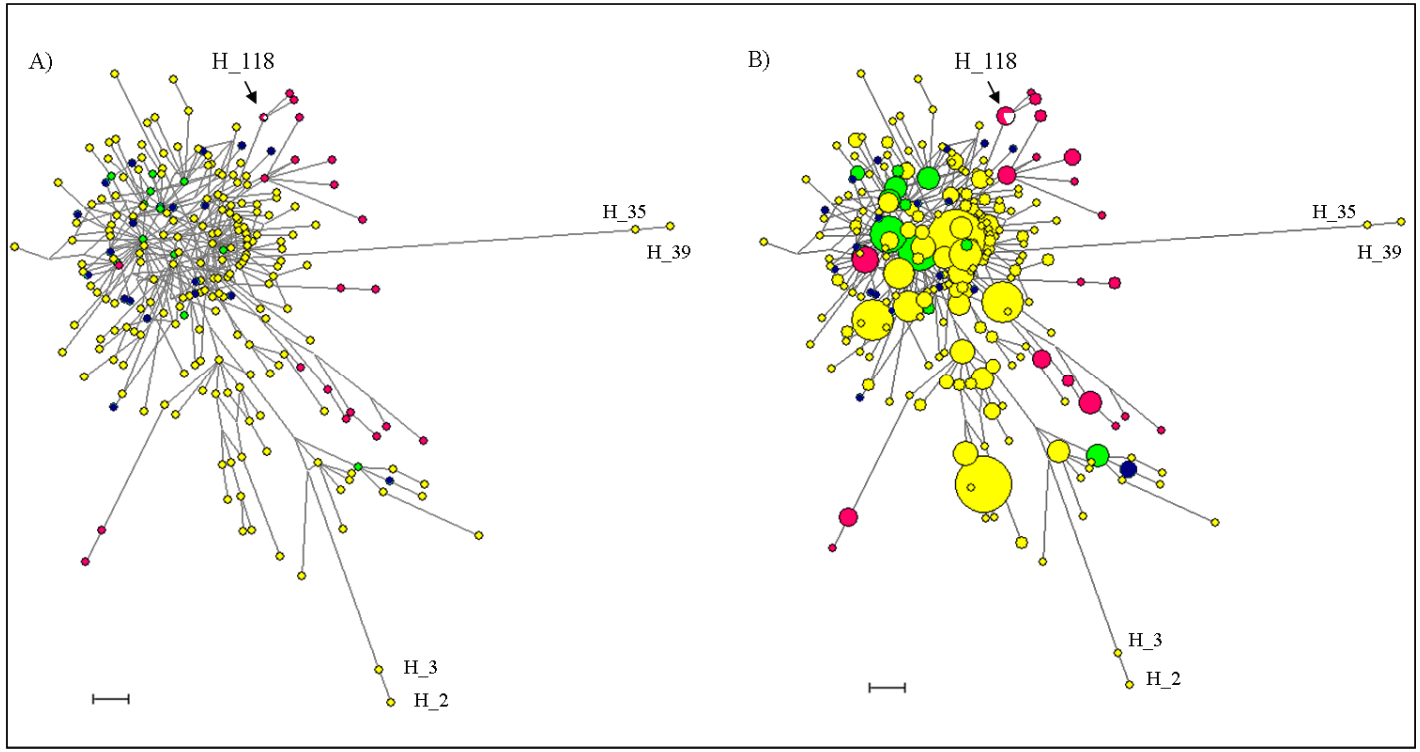


Figure 4. Mismatch distribution of mitochondrial DNA (CR + Cyt*b*) in the *agassizii* complex. Dashed line represents the observed mismatch distribution. Continuous line shows the expected distribution of the model under constant population size (A) and sudden expansion (B), according to the Roger's sudden expansion model (Rogers and Harpending 1992).

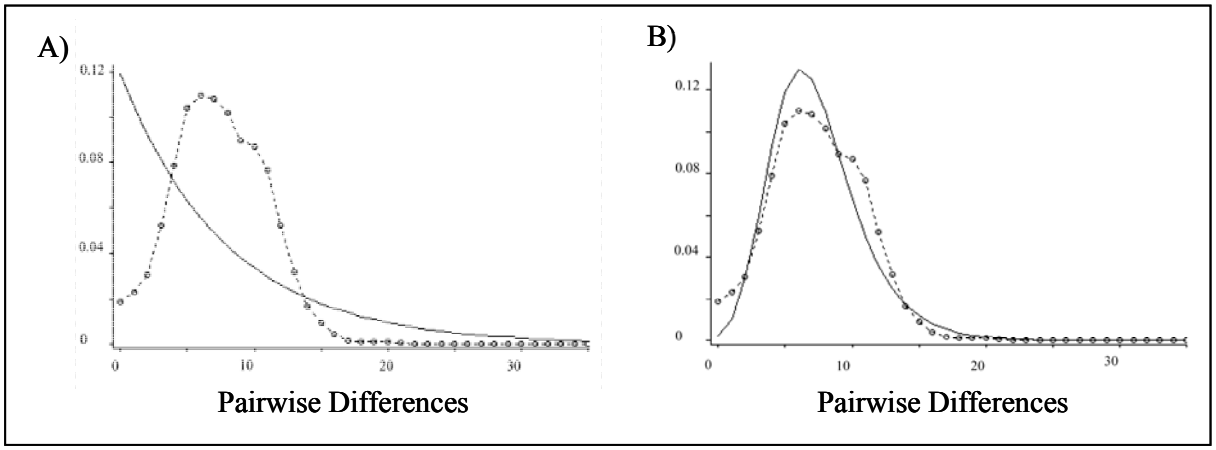


Figure 5. Plots of probabilistic assignments (*q*i) inferred from STRUCTURE for the *agassizii* complex across the Altiplano. Each specimen is represented by a vertical bar fragmented in *K* colored sections that is relative to its genetic proportion membership.

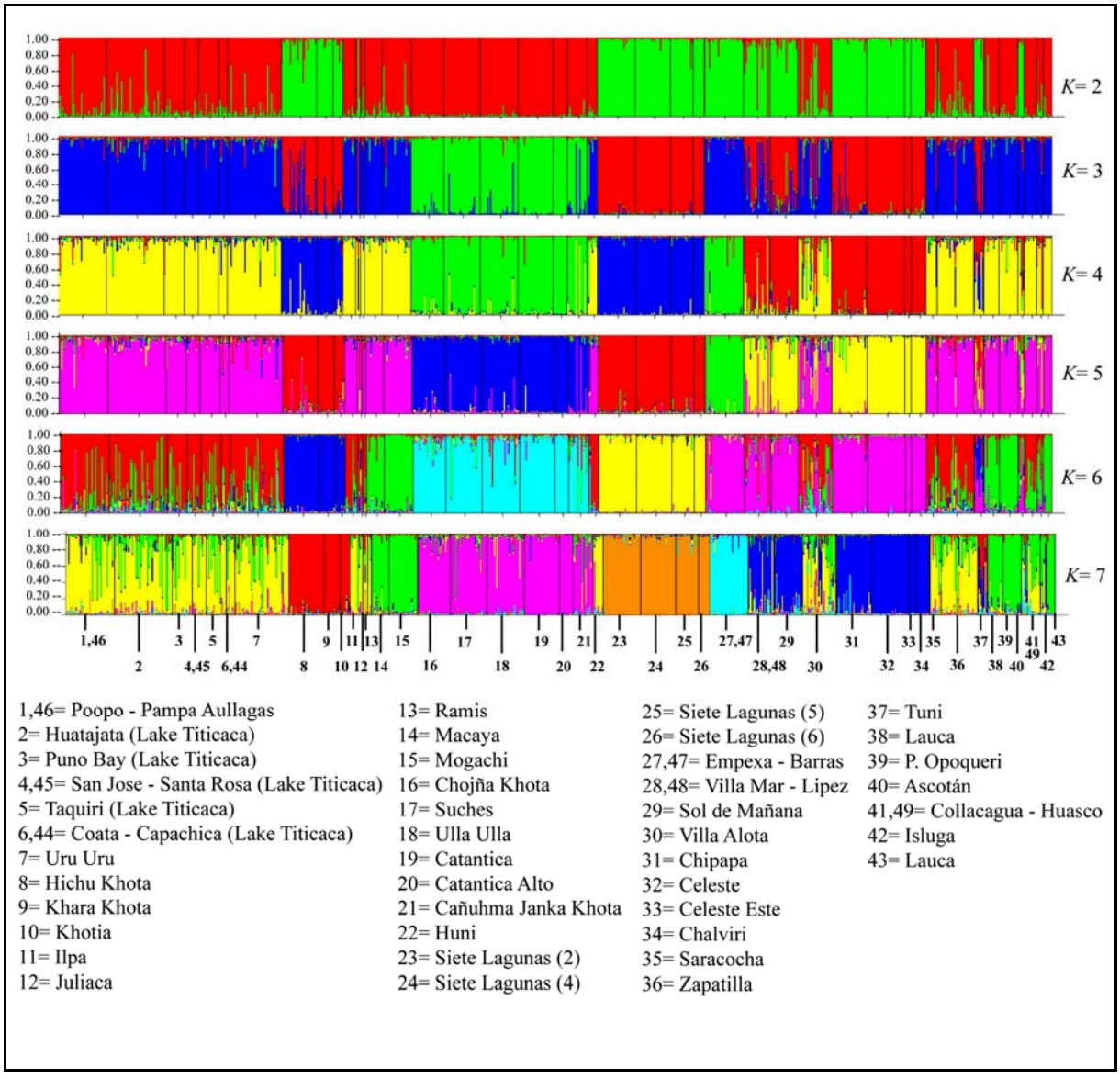


Figure 6. (see legend below)

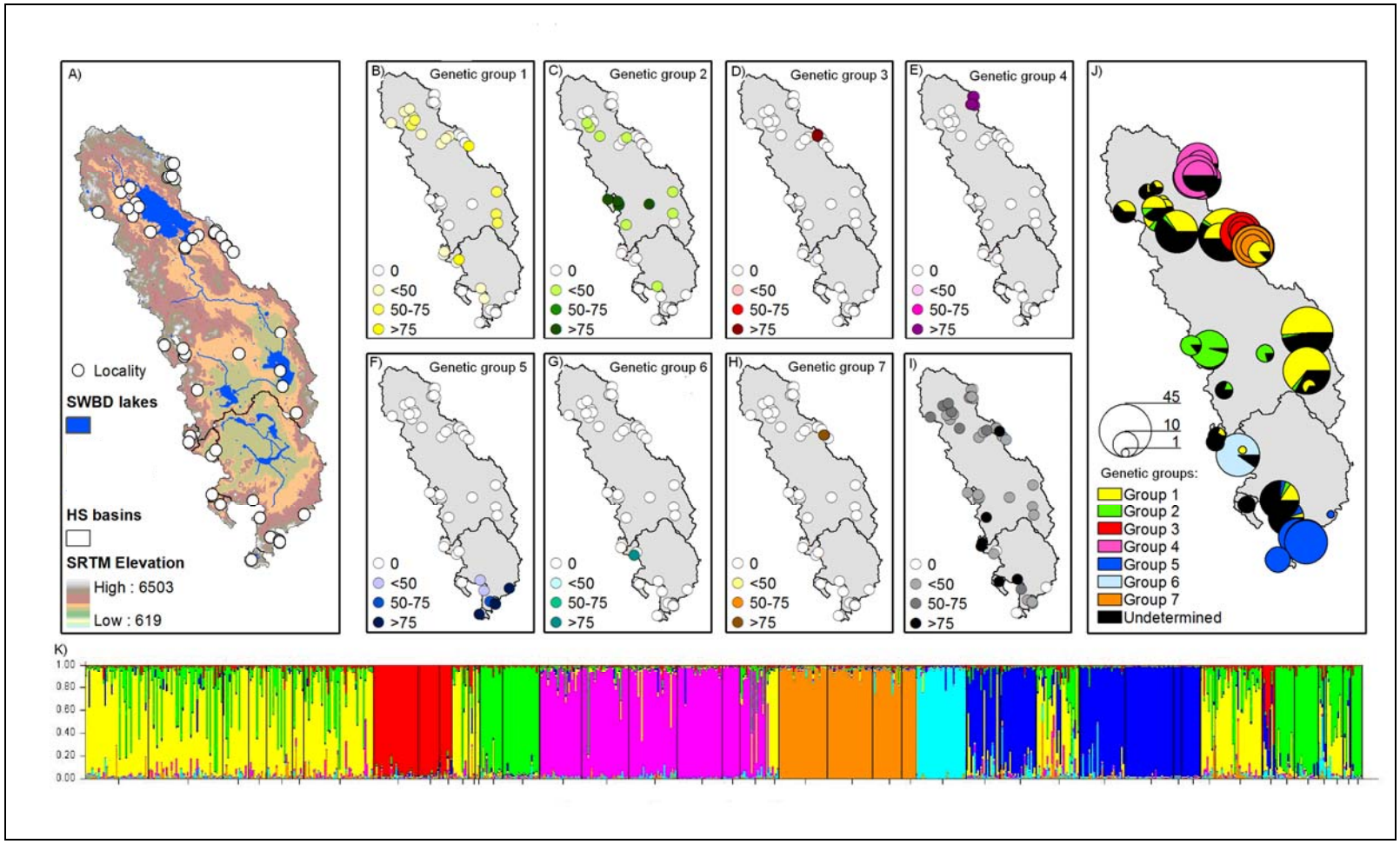


Figure 6. Geographic distribution of the genetic groups inferred in STRUCTURE (*K*=7) within the *agassizii* complex. Each locality is represented by a circle; circle color corresponds to the color assigned to each genetic cluster in the STRUCTURE plot (Fig. 5). (A) Localities replaced on an elevation map. The maximum extension of Lake Ballivian roughly follows the orange altitudinal delimitation (3650-3950m). (B to H) Geographic distribution for each genetic cluster, showing the proportion of individuals assigned to a specific cluster (when qi ≥ 0.80); color scale from light (0%) to dark (>75%). (I) Proportion of the specimens that could not be attributed to a genetic group as showing an assignment probability qi < 0.80. (J) Proportion of unambiguously assigned specimens (qi > 0.80; in color).

Figure 7. Mantel tests of correlation between geographic and genetic distances (F_{ST}) within the hydrographic group A using mtDNA sequences (A) and microsatellites (B).

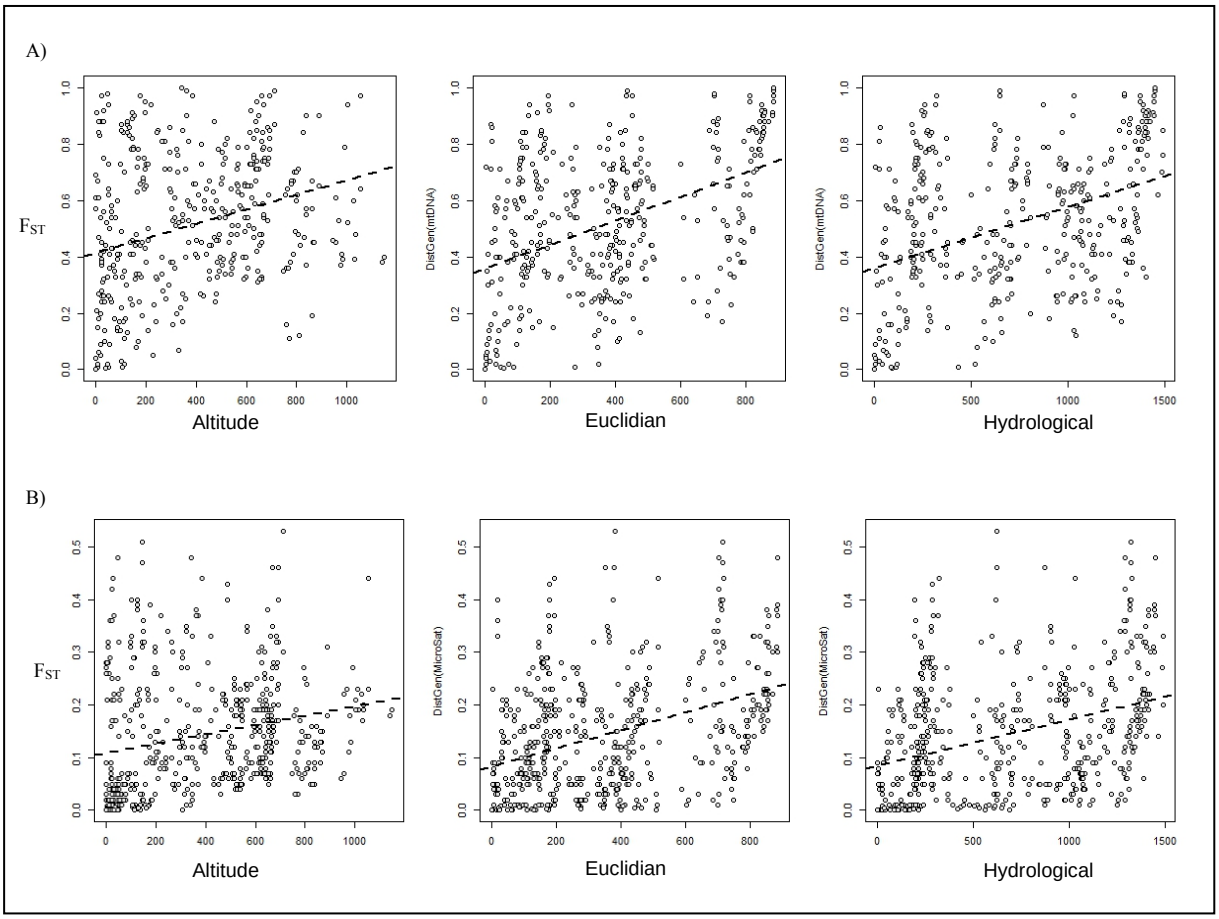


Figure 8. Relationships between geographic distances from Lake Titicaca and genetic distances (F_{ST}) of populations within the hydrographic group A, for mtDNA (CR + Cyt*b*) (A to C) and microsatellites (D to F).

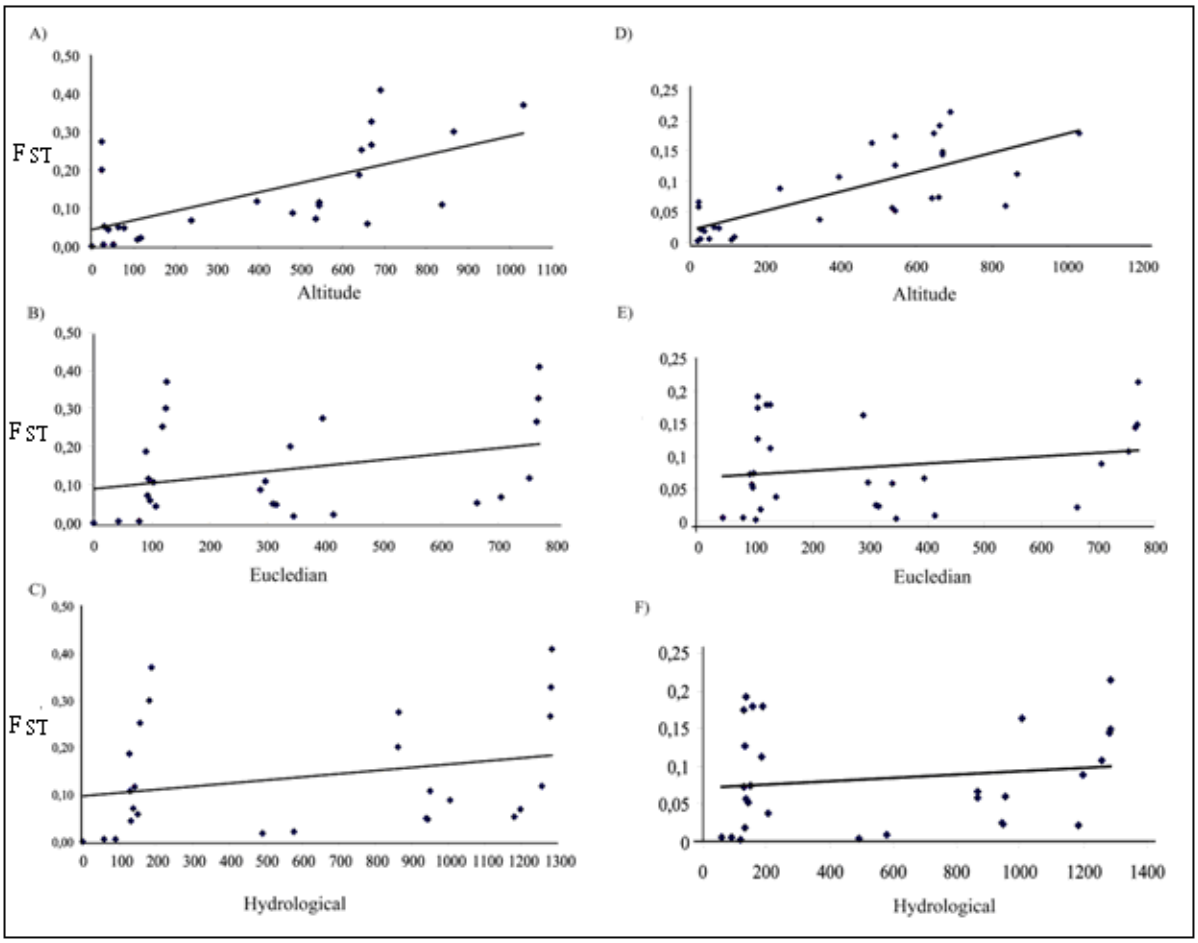
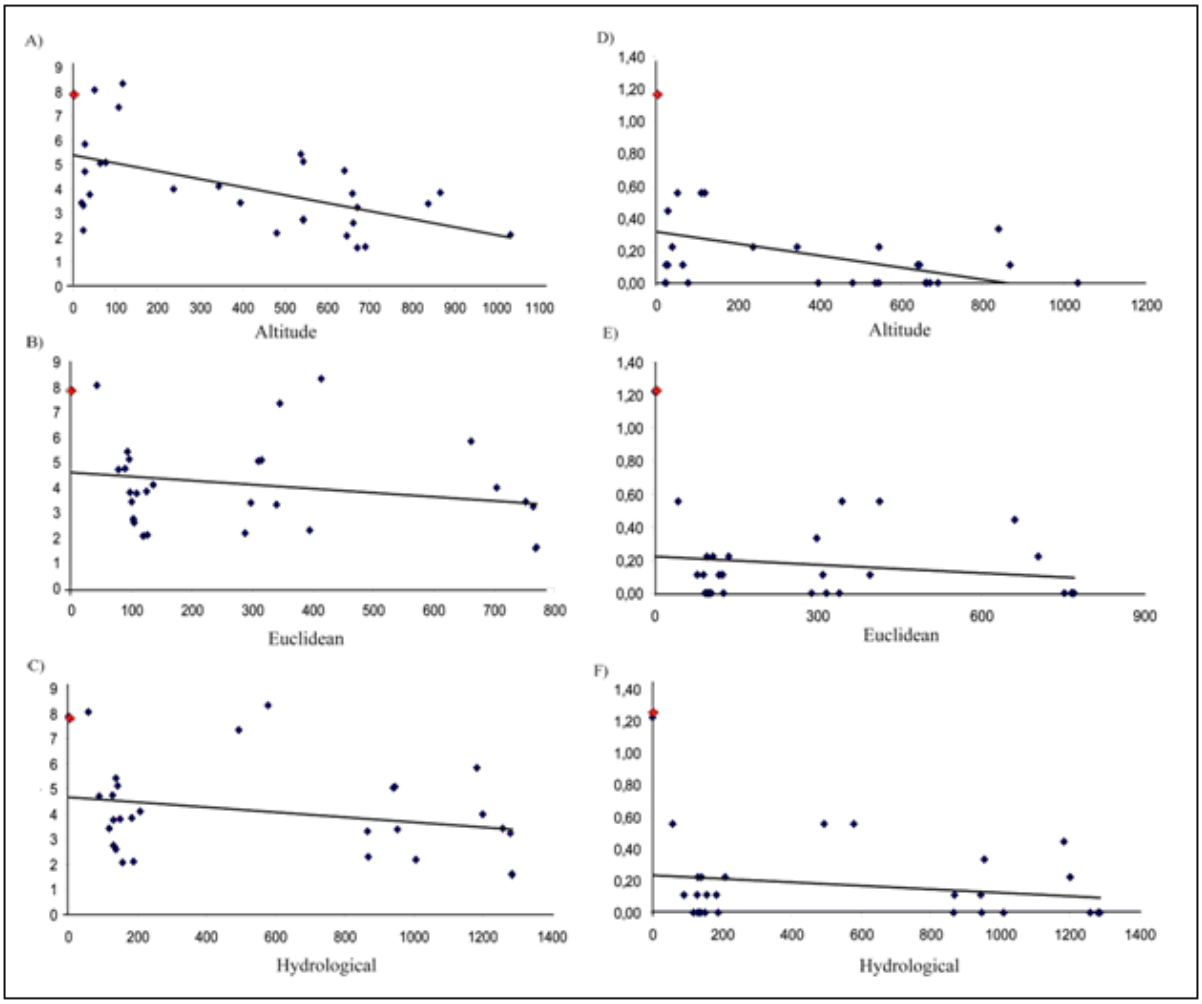


Figure 9. Relationships between geographic distances from Lake Titicaca and genetic diversity of populations within the hydrographic group A, measured as mean number of effective alleles N_E (A to C) and mean number of private alleles P_A (D to F), using microsatellites. Red diamond represents Lake Titicaca.



Tables.

Table 1. Demographic statistics estimated from mitochondrial DNA (CR + Cyt*b*) in the *agassizii* complex. Significance levels of observed statistics are specified as follows: *** ($P < 0.001$), ** ($0.001 < P < 0.01$), *($0.01 < P < 0.05$). SSD: Sum of squared deviations between observed and expected distributions, *r*: Harpending raggedness index. Statistics based on the frequency of mutations: *D**, *F** and *D* (Fu & Li 1993) and R_2 (Ramos-Onsins & Rozas 2002). Statistics based on haplotype distribution: *Fs*, Wall's *B* and *Q* (Wall 1999), Kelly's *ZnS* (Kelly 1997) and Roza's *ZA* (Rozas *et al*. 2001).

SSD	*r*	*D**	*F**	*Fs*	*D*	*ZnS*	*ZA*	*B*	*Q*	R_2
0.001	0.003 ***	- 5.411*	- 4.246*	- 362.90***	- 2.111**	0.006***	0.02	0.012	0.024	0.020**

Table 2. Mitochondrial DNA diversity estimates for the *agassizii* complex. Number of sequences (*N*), number of haplotypes (*h*), haplotype diversity (*Hd*), number of polymorphic sites (*S*) and nucleotide diversity per site (*Pi*).

	Partitions		
	CR	Cyt*b*	CR+ Cyt*b*
N	824	758	691
h	68	138	237
Hd	0.844	0.958	0.981
S	32	108	176
Pi	0.008	0.004	0.005

Table 3. Microsatellite diversity within the *agassizii* complex across the Altiplano. Locality identifier (ID), total number of samples (N), mean number of genotyped specimens (n), mean number of alleles (N_A), mean number of effective alleles (N_E), mean number of private alleles (P_A), mean fixation index (F_{IS}), and mean observed (H_O) and expected heterozygosity (H_E). Incidence of null alleles and Hardy Weinberg test are also given. Significance levels after Bonferroni correction for the Hardy Weinberg test (HW), as follows: ** ($0.001 < P < 0.01$), * ($0.01 < P < 0.05$).

Locality	ID	N	n	N_A	N_E	P_A	F_{IS}	H_O	H_E	Null Alleles	HW
Huatajata	2	46	43.6	13.444	7.046	0.222	0.018	0.653	0.685	A116, D110	** (A116)
Taquiri	5	16	15.11	10.667	6.091	0.222	0.004	0.664	0.717	A116, A106	** (A116)
Ilpa	11	10	9.33	6.556	4.697	0.111	-0.033	0.682	0.708		
Puno	3	16	15.11	10.111	6.489	0.444	-0.032	0.731	0.750		
Coata - Capachica	6 - 44	7	6.55	6.222	4.611	0.000	-0.044	0.667	0.705		
San Jose - Santa Rosa	4 - 45	5	5	6.000	5.051	0.222	0.122	0.667	0.847	D110, C102, A116	** (D110)
Juliaca	12	4	3.77	4.556	3.747	0.222	-0.089	0.704	0.755		
Ramis	13	3	2.77	3.889	3.420	0.000	-0.152	0.722	0.759		
Zapatilla	36	29	28	13.667	8.060	0.556	0.107	0.673	0.744	B103, A116, A106, C102	** (D110, B103)
Saracocha	35	8	6.22	5.444	5.272	0.222	-0.037	0.500	0.520		
Catantica	19	28	27.11	6.667	3.842	0.111	0.064	0.484	0.539		
Catantica Alto	20	11	10.88	2.667	2.107	0.000	-0.047	0.490	0.497		
Chojña Khota	16	26	24.66	9.111	4.762	0.111	0.137	0.484	0.598	A116, C105, A106, A9a	* (A116, C105, A106, A9a)
Suches	17	29	27.55	9.556	5.424	0.000	0.129	0.514	0.633	C105, A9a	** (C105, A9a)
Ulla Ulla	18	30	27.77	9.556	5.105	0.222	0.146	0.506	0.643	A116, C105, A9a	** (A116, B1, C105)
Cañuhma Janka Khota	21	16	14.66	6.222	3.792	0.000	0.099	0.488	0.566	C105, A116	* (D110, A116)
Hichu Khota	8	28	24	6.000	2.735	0.000	0.106	0.483	0.543	A116	** (A116)
Khotia	10	13	11.11	3.222	2.606	0.000	-0.018	0.505	0.501		
Khara Khota	9	8	7.33	3.889	2.711	0.000	0.036	0.517	0.538		
Tuni	37	8	7.77	2.778	2.068	0.111	0.037	0.306	0.347		
Siete Lagunas (2)	23	30	28.77	5.000	2.936	0.222	0.015	0.589	0.608		
Siete Lagunas (4)	24	28	25.44	4.333	2.836	0.111	0.009	0.581	0.597	A9a	** (A9a)
Siete Lagunas (5)	25	18	17.77	5.556	3.419	0.000	0.109	0.563	0.647		
Siete Lagunas (6)	26	9	8.33	3.667	2.633	0.000	-0.145	0.642	0.604		
Huni	22	8	6.22	2.778	2.498	0.111	0.258	0.333	0.431		

Table 3. (Continued)

Uru Uru	7	43	41.66	15.778	7.337	0.556	0.048	0.707	0.740	A116	
Poopo - Pampa Aullagas	1 - 4 6	39	36.77	14.778	8.330	0.556	0.033	0.696	0.743	A116	
P. Opoqueri	39	5	4.88	4.222	3.306	0.000	0.017	0.639	0.724		
Lauca (Bolivia)	38	12	12	7.667	5.102	0.000	-0.029	0.630	0.658		
Macaya	14	14	13.66	7.667	5.061	0.111	0.002	0.642	0.676	A106	** (A106)
Mogachi	15	23	22.22	6.222	3.390	0.333	0.184	0.528	0.606	C102, A116	** (B1)
Empexa - Barras	27 - 47	31	30.77	5.778	2.665	0.000	0.097	0.404	0.434	A116, B103, A106	* (D110)
Villa Mar - Sur Lipez	28 - 48	21	18	7.000	3.987	0.222	0.257	0.424	0.576	A116, C105, A9a	** (A116, C105)
Sol de Mañana	29	22	18.66	6.111	3.437	0.000	0.187	0.420	0.486	A116, C105	** (A116, C105)
Villa Alota	30	25	22.11	10.556	5.819	0.444	0.046	0.561	0.633	A116, C105	** (A116, B1)
Chipapa	31	28	24	5.556	3.648	0.000	0.005	0.324	0.357	A116	* (D110)
Celeste Este	33	30	26.33	3.444	1.821	0.000	-0.006	0.253	0.249	A116	
Celeste	32	5	4.22	2.000	1.793	0.000	0.104	0.237	0.309		
Chalviri	34	10	8.66	2.222	1.613	0.000	0.066	0.233	0.270		
Ascotán	40	5	4.77	4.111	3.122	0.000	0.057	0.541	0.655		
Collacagua	41	5	3.77	2.444	2.050	0.111	0.305	0.263	0.446		
Huasco	49	5	4.33	2.778	2.436	0.000	-0.133	0.417	0.426		
Isluga	42	5	4.11	2.778	2.300	0.111	-0.257	0.563	0.531		
Lauca (Chile)	43	7	6.55	3.111	2.175	0.000	0.308	0.302	0.405	B103	

Table 4. Results from Mantel tests of correlation (r) between geographic distances and F_{ST} (measured as Φ and θ for mtDNA and microsatellites, respectively).

Distance	Within hydrographic 'group A'		In the total study zone (hydrographic groups from A to G)	
	mtDNA (CR+Cyt*b*)	Microsatellites	mtDNA (CR+Cyt*b*)	Microsatellites
Altitude	0.220 (*P*=0.002)	0.191 (*P*=0.006)	0.190 (*P*=0.002)	0.161 (*P*=0.005)
Euclidian	0.380 (*P*=0.001)	0.363 (*P*=0.001)	0.343 (*P*=0.001)	0.313 (*P*=0.001)
Hydrological	0.331 (*P*=0.001)	0.330 (*P*=0.001)	NA	NA

Table 5. Correlation coefficients (r) between geographic and genetic distances and diversity from Lake Titicaca within the hydrographic group A.

	mtDNA	Microsatellites		
Distance	F_{ST} (Φ)	F_{ST} (θ)	N_E	P_A
Altitude	0.648 (*P*=0.000)	0.757 (*P*=0.000)	-0.295 (*P*=0.112)	-0.363 (*P*=0.048)
Euclidian	0.329 (*P*=0.099)	0.244 (*P*=0.193)	-0.173 (*P*=0.358)	-0.184 (*P*=0.329)
Hydrological	0.262 (*P*=0.195)	0.196 (*P*=0.297)	-0.244 (*P*=0.193)	-0.207 (*P*=0.270)

Electronic Material

Table S1. Pairwise genetic distances (Φ_{ST}) between localities calculated from mtDNA sequences (CR + Cyt*b*).

Table S2. Pairwise genetic distances (θ) between sampled localities calculated from microsatellite loci.

Chapter III

- **General Discussion and Perspectives:**

  the *agassizii* complex as a model of ongoing diversification?

III.1 Taxonomic implications

The delimitation of *Orestias* complexes and species has long been chaotic (Villwock 1986; Loubens 1989; Lauzanne 1992; Müller 1993; Villwock & Sienknecht 1995; Villwock & Sienknecht 1996) and is still in taxonomic chaos. However, before this investigation, molecular-based phylogenetic studies contrasting the traditional (morphological) systematics of the genus were lacking or very partial (Parker & Kornfield 1995; Lüssen *et al.* 2003). Our results combining ancient DNA-based extraction from type specimens and the taxonomic and geographically exhaustive collection of new samples contributed to clarify the phylogenetic delineation at the complex level within *Orestias* (Article 1). We redelimited four complexes, including *cuvieri, gilsoni, luteus* (here confirmed as a complex level group, i.e. regrouping a fair level of morphotypic diversity) and *agassizii* (which should exclude *O. jussiei*). The clade '*gilsoni*' is considered as provisionally named (on the basis of 'morphotypes' similar to described morpho-species belonging to the complex), since we could not include any related type-specimen in our analysis. In addition, we showed that *mulleri* was not a valid taxonomic entity (polyphyletic group). Remarkably, we could not recover the monophyly of any of the morpho-species considered in our sample set (including the extinct *O. cuvieri*), although we used one of the most fast-evolving mitochondrial genes in vertebrates (control region; similarly to Lüssen *et al.* 2003). We also provided evidence for overlapping ranges in diagnostic morphometric ratios previously used to distinguish between *O. agassizii* and *O. luteus,* likely owing to taxonomic malpractices (Article 3). As a consequence, our results severely question the validity of the diagnostic morphological characters traditionally considered in the definition of *Orestias* complexes and species, e.g. overall body shape and squamation pattern are the main criteria for defining *Orestias* species (see Lauzanne 1982; Parenti

1984b). It would thus appear interesting to engage into a taxonomic revision of the whole genus to redefine new diagnostic characters that would help taxonomists and field workers to identify Andean pupfishes with more certainty. Possible pathways could comprise (i) the use of our ancient DNA protocol of extraction (GuSCN buffer; Article 1) to include additional type series, and (ii) the use of more recent methods and technologies (morphometric geometrics, CT-scan) to extract new diagnostic morphological characters.

Our investigations allowed to further refine the delimitation of the *agassizii* complex (excluding the *luteus* group and *O. jussiei*) and to establish the latter as sister group of the three other complexes (Article 1). However, can we still speak of 'complex' if none of the morpho-species are found monophyletic? Incomplete lineage sorting has been posited as one of the main explanations for incongruence between gene trees and species trees and the non-recovery of monophyletic species, especially in groups having diversified in recent coalescent time (Pollard *et al.* 2006, Shaffer and Thomson 2007). Alternatively, but not exclusively, hybridization and backcrossing (asevidenced between *O. agassizii* and *O. luteus*; Article 3) have been proposed as events blurring species boundaries within recent diversification frameworks (Nyingi & Agnèse 2007), but also as potential promoters of speciation (Seehausen 2004). In order to clarify the processes at the origin of the unresolved phylogenetic pattern in the *agassizii* complex (but also in the other complexes), future efforts should concentrate on obtaining a large series of independent and variable markers —such as single nucleotide polymorphisms (SNPs)— possibly via the fast growing field of next generation sequencing (NGS) technologies (Wagner *et al.* 2013).

Last but not least, larger phylogenies including additional morpho-species (ECOS-CONICYT program; collaboration with Marco Mendez, Universidad de Chile)

and the northern range of the inter-Andean basin (Peru) might contribute to a better characterization of taxonomic delimitations within and of phylogenetic relationships among the *Orestias* complexes.

One central question remains open: how many *Orestias* species exist? Following specie's definition based under the lineage concepts (e.g. phylogenetic, evolutionary concept and the genealogical concepts), where monophyly is the centerpiece criterion (see dePinna 1999), we would be four species (i.e. one species per recovered clade/complex). This argument would fit with several of the first taxonomic revisions of the genus based on the observation of morphological characters, arguing that the species number within *Orestias* should be reduced (e.g. see Villwock 1986 for a discussion). This topic is still far for from being resolved and is in need of new analytical/technical approaches and empirical evidence.

III.2 A rapid diversification within the *agassizii* complex?

Our phylogenetic investigations on the genus *Orestias* revealed a 'deep' divergence between two main sister group lineages, one composed of the three complexes (mostly) endemic to Lake Titicaca (*cuvieri, gilsoni* and *luteus*) and one represented by *agassizii* inside and outside Lake Titicaca (Article 1). This divergence was then followed by more recent diversifications within each complex. These 'deep and shallow' diversification patterns would fit the Titicaca-centered adaptive radiation scenario of an 'ancient' divergence within Lake Titicaca between the two ancestors of the current sister group lineages, followed by ecological speciation with inter-specific competition and niche specialization within the lake during the Pleistocene, as posited by several authors (Parenti 1981; Parenti 1984b; Lauzanne 1982; Lüssen *et al.* 2003; Maldonado *et al.* 2009; Villwock 1986). Another remarkable result is that none of the morpho-species

considered either by Lüssen *et al*. (2003) or us (Article 1) could be found monophyletic. To the question "are we facing a genuine adaptive radiation within Lake Titicaca?", we are left without a definitive answer. Weak reproductive isolation barriers among morphologically divergent species as diagnosed between *O. agassizii* and *O. luteus* (Article 3) do not go against the framework of an ongoing adaptive radiation (Seehausen 2004). However, the first criterion of "rapid speciation rate, especially at early stage of diversification" enounced by Schluter (2000) to validate an adaptive radiation, cannot be properly tested given the absence of monophyletic species in *Orestias*. On the other hand, there is some support in the literature for the two other criteria given by the author, namely "adaptive nature of diversification, reflecting morphological, physiological and/or behavioral differences among species" and "ecophenotypes strongly correlated to biotic and abiotic characteristics of environment" (Loubens 1989; Lauzanne 1992; Northcote 2000; Maldonado *et al.* 2009). Thus, given the level of evidence available, one can hypothesize that the *Orestias* complexes within Lake Titicaca are undergoing an adaptive radiation, the coalescence time of which is too recent to have permitted the morpho-species to be isolated reproductively or at least to form monophyletic groups using the neutral genetic markers that we used (e.g. similar case in Tanganyika cichlids; Takahashi *et al.* 2001; Kocher 2004). This would match a scenario of recent speciation stage following Nossil's *et al*. (2009) criteria, characterized by i) low levels of reproductive isolation among taxa, ii) low to moderate levels of genotyping clustering (i.e. gene frequencies in individuals sampled from two taxa/populations), iii) weak level of lineage sorting among taxa (using both mitochondrial and nuclear DNA) and iv) low level of stepped geographical and/or ecological clines**.**

In a more general point of view, it is remarkable to notice that almost nothing is known about the behavior of Andean pupfishes and as a consequence, about the role of sexual selection in their lacustrine diversification (Takahashi & Koblmüller 2011; Wagner *et al.* 2012). The experimental design already set up for restocking programs (see §III.3) could be used to study further the sexual behavior of *Orestias* species.

Concerning the *agassizii* complex, our results tend to suggest that the species flock in Altiplano originated from the cyclic, maximum extension periods from paleolakes Ballivian to Tauca (0.6 Mya-14.5 kya) and subsequent dispersal into remote areas and basins, where populations were isolated following decrease in water levels (Article 4). Thus, paleolake fluctuations and connectivity among basins likely had a major impact on population structure and differentiation within the complex. Such a scenario of 'source-sinks' dynamics (Dias 1996) has already been posited in Cyprinodontiformes (see Barson *et al*. 2009). It is the first time that a hypothesis based on genetic investigations is provided to explain the diversification of the *agassizii* complex outside Lake Titicaca. We also showed a strong effect of isolation-by-distance across Altiplano, with an important effect of elevation (>200 m) on population differentiation from Lake Titicaca, despite regular flooding and intermittent connections among basins and populations. From the available evidence, we can posit a scenario of recent diversification within the *agassizii* complex that combines ecological speciation within Lake Titicaca (diversification with inter-specific competition) and allopatric speciation (mutation-based model) outside the lake (diversification without inter-specific competition, following long-distance dispersal). Thus, we could define the *agassizii* complex as a 'lacustro-riverine species flock', where both adaptive and non-adaptive radiations have shaped its diversification (see Rundell & Price 2009). We may hypothesize that speciation within Lake Titicaca has been driven by divergent natural

selection among ecological niches, whereas outside the lake mutation-order speciation occurred through mutations fixed in separate populations adapting to similar selection pressures (Schluter 2009). Nevertheless, contrary to the deep divergence times imposed by the non-adaptive radiation model (Rundell & Price 2009), diversification of the *agassizii* complex outside Lake Titicaca might be in its very early stage (e.g. extensive lineage sorting among mtDNA sequences). It would have been tempting to explain the genetic structure of the *agassizii* complex outside Lake Titicaca into several remote and isolated high-altitude populations by the surfing theory (Excoffier & Ray 2008). The latter predicts that low-frequency alleles surf on the wave of population range expansion, newly colonized areas becoming structured into populations of low genetic diversity distinguished by sharp allele frequency gradients. However, in our case, isolated populations did not show any particular pattern of low genetic diversity or high frequency of rare alleles. This suggests that numbers of founders were not low (possibly due to secondary dispersals) (Roman & Darling 2007) and/or that some level of connectivity within isolated populations (through migrants) allows for maintaining genetic diversity (Lacy 1987). In terms of perspectives, explaining how those Andean pupfishes with poor swimming abilities may disperse and connect through large altitudinal and ecological gradients and unsuitable habitats across the Altiplano will require additional investigations on their life-history traits, which are almost unknown.

The hypothesized adaptive radiation within Lake Titicaca (*cuvieri, gilsoni, luteus* and part of *agassizii*) and non-adaptive radiation across the Altiplano (rest of *agassizii*) yielded in a seemingly similar period of time (Article 1) an almost equivalent number of morpho-species (22 species *versus* 17 species, respectively). This supports the idea that both sympatric adaptive radiation (McCune & Lovejoy 1998) and allopatric non-adaptive radiation (related to niche conservatism; Near & Michael 2004)

can result in high speciation rates. However, in the *agassizii* complex, remains to be assessed to what extent the morpho-species present in the lake are eco-morphologically separated, and on the contrary are ecologically similar outside the lake. Ecological and morphological investigations should be conducted to redefine the limits of the morpho-species and the association of their functional morphology with habitat and trophic niche. Disentangling between the nature of adaptive vs. non-adaptive radiation within the *agassizii* complex (and more generally, within *Orestias*) should benefit from NGS-based investigations capable of generating a great amount of independent and variable markers (e.g. SNPs) across large portions of the genome. In such a context, the ongoing assembling of the first genome draft of *Orestias agassizii* (Marco Méndez, Universidad de Chile) will deliver exciting opportunities to investigate the genomic landscape of Andean pupfishes and to identify genes under selection.

On a more general perspective, a divergence time framework for the diversification of the genus has been cruelly lacking. The absence of fossils representing *Orestias* or its sister-group (Parker & Kornfield 1995), but see Costa (2012), renders difficult to accurately calibrate the phylogeny of the genus (Lüssen *et al.* 2003). One solution would be to rely on secondary calibration points from deeper phylogenies or to increase the number of loci for which evolutionary rates among teleosteans are relatively well-known, but none of those options are optimal (Blair Hedges & Kumar 2004; Mueller 2006).

III.3 Conservation implications

Several studies have shown that the relative rate of extinction of freshwater species is comparable to that of species in tropical forest (Ricciardi & Rasmussen 1999). The situation is especially alarming in semi-arid and arid regions such as the Altiplano,

where freshwater habitats are fragile and demand for water resources is growing (Vila *et al.* 2007b). Habitat deterioration includes overfishing, sediment loading and organic pollution from land use activities (e.g. mining), industrial toxic contaminants, damming and drainage of waterways, and release of invasive species.(Lauzanne 1992; Northcote 1992; Orlove *et al.* 1992; Vila *et al.* 2007a; Vila *et al.* 2007b).

The recent development of conservation genetics has provided valuable insights into the regional-scale evolution of freshwater teleosteans that can be used for conservation management (Raeymaekers *et al.* 2009; Strecker *et al.* 2011). At a larger scale, our phylogenetic redefinition of the *Orestias* complexes (Article 1) identified *cuvieri* as an extinction-prone clade composed of specialized morphotypes (McKinney 1997), including *O. cuvieri* (large, pelagic and piscivorous; extinct), *O. pentlandii* (large, pelagic and feeding on zooplankton, copedoda and cladocera; almost not found in the wild), *forgeti* (small, pelagic and feeding on zooplankton, copedoda and cladocera; almost absent in the wild) and *ispi* (small, pelagic and feeding on zooplankton, copedoda and cladocera; the only common representative) (Harrison & Stiassny 2004 http://creo.amnh.org/pdi.html; Van Damme *et al.* 2009).

The most straightforward implication of our study is the identification across the central and southern Altiplano —using the diagnostic series of microsatellite loci that we developed (Article 2)— of seven genetic populations within the *agassizii* complex that likely constitute distinct evolutionary significant units (ESUs; Moritz 1994) or species (Article 4). Thus, our investigations shed a new light on lineage delimitation within the *agassizii* complex, which has been traditionally based on morphological characters (inapplicable criteria; see above). Several ESUs were isolated populations in restricted and fragile high-altitude habitats, including Sur Lipez, Ulla Ulla, Hichu Khota valley and Siete Lagunas. Those represent typical cases of “island populations”, which

are more likely to be subject to extinction proneness due to their demographic characteristics, notably via inbreeding (Frankham 1998; Houde *et al.* 2011). However, given the heterogeneous level of F_{IS} detected within the isolated ESUs, inbreeding may not represent a significant risk (extinction) at the current stage (see Article 4). During my PhD, we could witness ongoing processes of extinction in one of the isolated populations (Siete Lagunas), with four out of seven of the small lakes having dried out in less than 4 years. Together with the evidence for outbreeding (F_{IS}) within these restricted ESUs, this raises the question of how such isolated populations may survive and maintain a certain level of genetic diversity. Micro-scale population genetic studies will have to be conducted to accurately assess the potential connectivity pathways that could exist within each ESU.

The identification of several ESUs across the Altiplano let us expect the delimitation of additional ESUs from the not yet sampled area in the northern part of the inter-Andean basin (Peru), where most of the morpho-species of the *agassizii* complex have been described. On the other hand, our analyses failed to recover distinct genetic populations for any of the Chilean endemic morpho-species considered (*O. ascotanensis*, *O. gloriae*, *O. laucaensis* and *O. sp* from Rio Lauca; see Article 4), despite observed variation in morphology and karyotypes (Parenti 1984b; Vila *et al.* 2010). Although our results are preliminary, the issue of whether or not those hypothesized morpho-species —considered of high conservation concern for the country (Harrison & Stiassny 2004; Van Damme *et al.* 2009)— are ecotypes (e.g. described ecotypes in: Lauzanne 1992; Northcote 2000; Maldonado *et al.* 2009) should be rigorously tested.

Finally, the accurate level of genetic delimitation reached within the *agassizii* complex (ESUs; Article 4) and for hybrids (*O. agassizii x O. luteus;* Article 3), should

constitute a useful basis to orientate the artificial spawning programs and restocking of *Orestias* species. In Peru, such programs have traditionally been focused on the most appreciated species in local fisheries, including *O. luteus, O. agassizii*, *O. ispi*, *O. albus* and *O. pentlandii* (Treviño *et al.* 1992; Castañon Rivera 1994; Castañon Rivera *et al.* 1995; Polo Sanchez 2005). The genetic diagnostic tools that our investigations have managed to provide should be used to develop new programs on extinction prone species and populations(Rodriguez-Ramilo *et al.* 2009) while avoiding the risks of cross-species hybridization and outbreeding depression , notably among Lake Titicaca and the genetically differentiated populations of the *agassizii* complex (Article 4). The release of hybrid specimens produced under laboratory conditions —a practice that seems well established to increase survival or growth rate; e.g. *O. agasizii* x *O. pentlandii*, *O. luteus* x *O. agassizii* (Aspiazu 2002)— should be avoided in order to ensure the genetic cohesion of wild species. On the other hand, the natural hybrids with slightly different morphologies than intermediary between *O. agassizii* and *O. luteus* that we detected in Lago Pequeño (Lake Titicaca) and Lake UruUru (Article 3) may be considered as potential ongoing sources of diversification (hybrid swarm theory; Seehausen 2004) that should also receive attention from conservation programs.

References

Albrecht, C. & Wilke, T. (2008) Ancient Lake Ohrid: biodiversity and evolution. *Hydrobiologia*, 615, 103-140.

Argollo, J. & Mourguiart, P. (2000) Late Quaternary climate history of the Bolivian Altiplano. *Quaternary International*, 72, 37-51.

Aspiazu, W.I. (2002) Hibridación artificial entre carachi (*Orestias agassii*) y punku (*Orestias luteus*) "killifish" del Lago Titicaca. *In*. Facultad de Agronomía, Universidad Mayor de San Andrés, Tésis de Licenciatura, La Paz, pp. 62.

Barson, N.J., Cable, J., Van Oosterhout, C. (2009) Population genetic analysis of microsatellite variation of guppies (*Poecilia reticulata*) in Trinidad and Tobago: evidence for a dynamic source-sink metapopulation structure, founder events and population bottlenecks. *Evolutionary Biology*, 22, 485-497.

Berra, T.M. (2001) *Freshwater fish distribution*. Academic Press, San Diego, 604 pp.

Blair Hedges, S. & Kumar, S. (2004) Precision of molecular time estimates. *Trends in Genetics*, 20, 242-247.

Blodgett, T.A., Isacks, B.L. & Lenters, J.D. (1997) Constraints on the origin of paleolake expansions in the central Andes. *Earth Interactions*, 1, 1-28.

Capriles, J.M., Domic, A.I. & Moore, K.M. (2008) Fish remains from the formative period (1000 BC-AD 400) of Lke Titicaca, Bolivia: zooarcheology and taphonomy. *Quaternary International*, 180, 115-126.

Carmouze, J.P., Arce, C. & Quintanilla, J. (1981) Régulation hydrochimique du lac Titicaca et l'hydrochimie de ses tributaires. *Revue d'Hydrobiologie Tropicale*, 14, 329-348.

Carson, E. & Dowling, E. (2006) Influence of hydrogeographic history and hybridization on the distribution of genetic variation in the pupfishes *Cyprinodon atrorus* and *C. bifasciatus*. *Molecular Ecology*, 15, 667-679.

Castañon Rivera, V.A. (1994) Evaluación de técnicas de desove e incubación artificial para *Orestias agassii* y *Orestias luteus*. *In*. Facultad de Agronomía, Universidad Mayor de San Andrés, Tésis de Licenciatura, La Paz, pp. 130.

Castañon Rivera, V.A., De la Quintana Gonzales, H. & Limachi Corani, J. (1995) Reproduccion artificial de ispi (*Orestias ispi*). *In:* Manual Técnico. Centro de Investigacion y Desarrollo piscicola del Altiplano (CIDPA), La Paz, Bolivia

Chen, W.-J., Bonillo, C. & Lecointre, G. (2003) Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. *Molecular Phylogenetics and Evolution*, 26, 262-288.

Chow, S. & Hazama, K. (1998) Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology*, 7, 1247-1263.

Costa, W.J.E.M. (1997) Phylogeny and classification of the Cyprinodontidae revisited (Teleostei: Cyprinodontiformes): are Andean and Anatolian killifishes sister taxa? *Journal of Computational Biology*, 2, 1-17.

Costa, W.J.E.M. (2012) Redescription and phylogenetic position of the fossil killifish † *Carrionellus diumortuus* White from the Lower Miocene of Ecuador (Teleostei: Cyprinodontiformes). *Cybium*, 35, 181-187.

Cristescu, M.E., Adamowicz, S.J., Vaillant, J.J. & Haffner, D.G. (2010) Ancient lakes revisited: from the ecology to the genetics of speciation. *Molecular Ecology*, 19, 4837-4851.

Cross, S.L., Baker, P.A., Seltzer, G.O., Fritz, S.C. & Dunbar, R.B. (2001) Late Quaternary climate and hydrology of tropical south America inferred from an

isotopic and chemical model of lake Titicaca, Bolivia and Peru. *Quaternary Research*, 56, 1-9.

Cuvier, G. & Valenciennes, A. (1846) *Histoire des Poissons*. Paris, 505 pp.

Danley, P.D., Husemann, M., Ding, B., DiPietro, L.M., Beverly, E.J. & Peppe, D.J. (2012) The impact of the geologic history and paleoclimate on the diversification of East African cichlids. *International Journal of Evolutionary Biology*, 2012, 20.

Danley, P.D. & Kocher, T.D. (2001) Speciation in rapidly diverging systems: lessons from Lake Malawi. *Molecular Ecology*, 10, 1075-1086.

de Graaf, M., Megens, H.-J., Samallo, J. & Sibbing, F. (2010) Preliminary insight into the age and origin of the *Labeobarbus* fish species flock from Lake Tana (Ethiopia) using the mtDNA cytochrome *b* gene. *Molecular Phylogenetics and Evolution*, 54, 336-343.

Decou, A., von Eynatten, H., Dunki, I., Frei, D. & Wörner, G. (2013) Late Eocene to Early Miocene Andean uplift inferred from detrital zircon fission track and U-Pb dating of Cenozoic forearc sediments (15-18°S). *Journal of South American Earth Sciences*, 45, 6-23.

Dejoux, C. (1992) The Tricladida and Hirudinae. *In:* Dejoux, C. & Iltis, A. (Eds.) *Lake Titicaca: a synthesis of limnological knowledge*. Kluwer Academic Publishers, Dordrecht, pp. 307-310.

de Pinna, M.C.C. (1999) Species concepts and phylogenetics. *Reviews in Fish Biology and Fisheries*, 9, 353-373.

Dias, P.C. (1996) Sources and sinks in population biology. *Trends in Ecology & Evolution*, 11, 326-330.

Duda, T. F. & Rolán E. (2005) Explosive radiation of Cape Verde *Conus,* a marine species flock. *Molecular Ecology*, 14, 267-272

Eastman J.T. & McCune A.R. (2000) Fishes on the Antarctic continetal shelf: evolution of a marine species? *Journal of Fish Biology*, 57, 84-102.

Echelle, A.A. & Kornfield, I. (1984) *Evolution of fish species flocks*. University of Maine Press, Orono, 257 pp pp.

Eigenmann, C. (1910) Catalogue of the freshwater fishes of tropical and south temperate America. *In:* Scott, W. (Ed.) *The Princeton University Expeditions to Patagonia, 1896-1899*. Princeton University, Princeton, pp. 375-511.

Eigenmann, C.H. (1920) On the genera *Orestias* and *Empetrichthys*. *Copeia*, 89, 103-106.

Eschmeyer, W. & Fong, J. (2013) Species of fishes by family/subfamily. *In:* Online catalogue of fishes http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp . Mai 2013.

Excoffier, L. & Ray, N. (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution*, 23, 347-351.

Falk, T.M., Teugels, G.G., Abban, E.K., Villwock, W. & Renwrantz, L. (2003) Phylogeographic patterns in populations of the black-chinned tilapia complex (Teleostei, Cichlidae) from coastal areas in West Africa: support for the refuge zone theory. *Molecular Phylogenetics and Evolution*, 27, 81-92.

Fornari, M., Risacher, F. & Féraud, G. (2001) Dating of paleolakes in the central Altiplano of Bolivia. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 172, 269-282.

Foster, N.P. (1967) Trends in the evolution of reproductive behavior in killifishes. *Studies in Tropical Oceanography*, 5, 549-566.

Frankham, R. (1998) Inbreeding and extinction: island populations. *Conservation Biology*, 12, 665-675.

Garman, S.W. (1876) Exploration of lake Titicaca by A. Agassiz and S.W. Garman. *Bulletin of the Museum of Comparative Zoology, Harvard University*, 3, 274-278.

Garman, S.W. (1895) *The cyprinodonts*. Memoirs of the Museum of Comparative Zoology, Cambridge, 1-179 pp.

Ghosh, P., Garzione, C.N. & Eiler, J.M. (2006) Rapid uplift of the Altiplano revealed through 13C-18O bonds in paleosol carbonates. *Science*, 311, 511-515.

Gillespie, R. (2004) Community assembly through adaptive radiation in Hawaiian spiders. *Science*, 303, 356-359.

Gittenberger, E. (1991) What about non-adaptive radiation? *Biological Journal of the Linnean Society*, 43, 263-272.

Givnish, T.J. (1997) Adaptive radiation and molecular systematics: issues and approaches. *In:* Givnish, T.J. & Sytsma, K.J. (Eds.) *Molecular Evolution and Adaptive Radiation*. Cambridge University Press, Cambridge, UK, pp. 1-54.

Gonzalez, E.R. & Watling, L. (2003) A new species of *Hyalella* from the Patagonia, Chile with re-description of *H. simplex* Schellenberg, 1943 (Crustacea: Amphipoda). *Journal of Natural History*, 37, 2077-2094.

Gorthner, A. (1994) What is an ancient lake? *In:* Martens, K., Goddeeries, B. & Coulter, G. (Eds.) *Speciation in ancient lakes*. Archiv für Hydrobiologie, pp. 97-100.

Greenwood, P.H. (1984) What is a species flock? . *In:* Echelle, A.A. & Kornfield, I. (Eds.) *Evolution of fish species flock*. University of Maine, Orono, pp. 13-19.

Günther, A. (1866) A catalogue of the fishes in the British Museum. Vol. 6, London

Harrison, I.J. & Stiassny, M.L.J. (2004) CREO list of fish extinctions since AD 1500. American Museum of Natural History, Committee on Recently Extinct Organism.

Hendry, A., P, Nosil, P. & Rieseberg, L.H. (2007) The speed of ecological speciation. *Functional Ecology*, 21, 455-464.

Hershler, R. & Thompson, F.G. (1992) A review of the aquatic gastropod subfamily Cochliopinae (Prosobranchia:Hydrobiidae). *Malacological Review Supplement*, 5, 1-140.

Hoorn, C., Wesselingh, F.P., ter Steege, H., Bermudez, M.A., Mora, A., Sevink, J., Sanmartín, I., Sanchez-Meseguer, A., Anderson, C.L., Figueiredo, J.P., Jaramillo, C., Riff, D., Negri, F.R., Hooghiemstra, H., Lundberg, J., Stadler, T., Särkinen, T. & Antonelli, A. (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, 330, 927-931.

Houde, A.L.S., Fraser, D.J., O'Reilly, P. & Hutchings, J.A. (2011) Relative risks of inbreeding and outbreeding depression in the wild in endangered salmon. *Evolutionary Applications*, 4, 634-647.

Hulsey, C., Mims, M., Parnell, N. & Streelman, J. (2010) Comparative rates of lower jaw diversification in cichlid adaptive radiations. *Journal of Evolutionary Biology*, 23, 1456-1467.

Johns, G.C. & Avise, J.C. (1998) Test for ancient species flocks based on molecular phylogenetic appraisals of *Sebastes* rockfishes and other marine fishes. *Evolution*, 52, 1135-1145.

Johnson, T.C., Scholz, C.A., Talbot, M.R., Kelts, K., Ricketts, D., Ngobi, G., Beuning, K., Ssemmanda, I. & McGill, J.W. (1996) Late Plesitocene desiccation of lake Victoria and rapid evolution of cichlid fishes. *Science*, 273, 1091-1093.

Keller, B. & Soto, D. (1998) Hydrogeologic influences on the preservation of *Orestias ascotanensis* (Teleostei: Cyprinodontidae), in Salar de Ascotan, northern Chile. *Revista Chilena de Historia Natural*, 71, 147-156.

Kocher, T.D. (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics*, 5, 288-298.

Kosswig, C. & Villwock, W. (1964) Das problem der intralakustrischen speziation im Titicaca und Lanaosee. *Verhandlungen der Deutschen Zoologischen Gesellschaft*, 28, 95-102.

Kozak, K.H., Weisrock, D.W. & Larson, A. (2006) Rapid lineage accumulation in a non-adaptive radiation: phylogenetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: *Plethodon*). *Proceedings of the Royal Society B: Biological Sciences*, 273, 539-546.

Kroll, O., Hershler, R., Albrecht, C., Terrazas, E.M., Apaza, R., Fuentealba, C., Wolff, C. & Wilke, T. (2012) The endemic gastropod fauna of lake Titicaca: correlation between molecular evolution and hydrographic history. *Ecology and Evolution*, 2, 1517-1530.

Kundzewicz, Z.W., Mata, L.J., Arnell, N.W., Döll, P., Jimenez, B., Miller, K., Oki, T., Z., Ş. & Shiklomanov, I. (2008) The implications of projected climate change for freshwater resources and their management. *Hydrological Sciences Journal*, 53, 3-10.

Lacy, R.C. (1987) Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conservation Biology*, 1, 143-158.

Lauzanne, L. (1982) Les *Orestias* (Pisces, Cyprinodontidae) du petit lac Titicaca. *Revue d'Hydrobiologie Tropicale*, 15, 39-70.

Lauzanne, L. (1992) Native species. The *Orestias*. *In:* Dejoux, C. & Iltis, A. (Eds.) *Lake Titicaca: a synthesis of limnological knowledge*. Kluwer Academic Publishers, Dordrecht, pp. 405-419.

Lavenu, A. (1992) Formation and geological evolution. *In:* Dejoux, C. & Iltis, A. (Eds.) *Lake Titicaca: a synthesis of limnological knowledge*. Kluwer Academic Publishers, Dordrecht, pp. 3-15.

Lavenu, A., Fornari, M. & Sebrier, M. (1984) Existence de deux nouveaux épisodes lacustres quaternaires dans l'Altiplano péruvo-bolivien. *Cahiers ORSTROM série Géologie*, 14, 103-114.

Li, C., Orti, G., Zhang, G. & Guoqing, L. (2007) A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evolutionary Biology*, 7, 1-11.

Losos, J. (2010) Adaptive radiation, ecological opportunity, and evolutionary determinism. *The American Naturalist*, 175, 623-639.

Loubens, G. (1989) Observations sur les poissons de la partie Bolivienne du lac Titicaca. 4. *Orestias* spp., *Salmo gairdneri* et problémes d'aménagement. *Revue d'Hydrobiologie Tropicale*, 22, 157-177.

Lüssen, A., Falk, T. & Villwock, W. (2003) Phylogenetic patterns in populations of Chilean species of the genus *Orestias* (Teleostei: Cyprinodontidae): results of mitochondrial DNA analysis. *Molecular Biology and Evolution*, 29, 151-160.

Machordom, A. & Doadrio, I. (2001) Evidence of a Cenozoic Betic–Kabilian connection based on freshwater fish phylogeography (*Luciobarbus*, Cyprinidae). *Molecular Phylogenetics and Evolution*, 18, 252-263.

Maldonado, E., Hubert, N., Sagnes, P. & De Mérona, B. (2009) Morphology-diet relationships in four killifishes (Teleostei, Cyprinodontidae, *Orestias*) from Lake Titicaca. *Journal of Fish Biology*, 74, 502-520.

Martens, K. (1997) Speciation in ancient lakes. *Trends in Ecology & Evolution*, 12, 177-182.

Mayr, E. (1942) *Systematics and the origin of species* Columbia University Press, New York, 334 pp.

Mayr, E. (1984) Evolution of fish species flocks: a commentary. *In:* Echelle, A.A. & Kornfield, I. (Eds.) *Evolution of fish species flocks* University of Maine, Orono, pp. 21-25.

McCartney, M.A., Acevedo, J., Heredia, C. *et al*. (2003) Genetic mosaic in a marine species flock. *Molecular Ecology*, 12, 2963-2973.

McCune, A.R. & Lovejoy, N.R. (1998) The relative rate of sympatric and allopatric speciation in fishes. *In:* Howard, D.J. & Berlocher, S.H. (Eds.) *Endless forms: species and speciation*. Oxford University, New York, pp. 172-185.

McKinney, M.L. (1997) Extinction vulnerability and selectivity: combining ecological and paleontological views. *Ecology, Evolution and Systematics*, 28, 495-516.

Morales, P., Vila, I. & Poulin, E. (2011) Genetic structure in remnant populations of an endangered cyprinodotid fish, *Orestias ascotanensis*, endemic to the Ascotan salt pan of the Altiplano. *Conservation Genetics*, 12, 1639-1643.

Moritz, C. (1994) Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology & Evolution*, 9, 373-375.

Mueller, R.L. (2006) Evolutionary rates, divergence dates, and the performance of mitochondrial genes in bayesian phylogenetic analysis. *Systematic Biology*, 55, 289-300.

Müller, R. (1993) Critical remarks on the revision of the genus *Orestias* (Pisces Cyprinodontidae) by Parenti (1984). *Zoologische Jahrbuecher Abteilung fuer Anatomie und Ontogenie der Tiere*, 123, 31-58.

Myers, G.S. (1931) The primary groups of oviparous cyprinodont fishes. *Standford University Publications: Biological Sciences*, 6, 1-14.

Nagle, B.C. & Simons, A.M. (2012) Rapid diversification in the North American minnow genus *Nocomis*. *Molecular Phylogenetics and Evolution*, 63, 639-649.

Near, T.J. & Michael, F.B. (2004) Rapid allopatric speciation in longperch darters (Percidae: *Percina*). *Evolution*, 28, 2798-2808.

Nelson, J.S. (2006) *Fishes of the World*. John Willey & Sons, New Jersey, 642 pp.

Niemeyer, H. & Cereceda, P. (1984) Hidrografía. *In: Geografía de Chile*. Instituto Geográfico Militar, Santiago, Chile, pp. 300.

Northcote, T.G. (1992) Eutrophication and pollution problems. *In:* Dejoux, C. & Iltis, A. (Eds.) *Lake Titicaca: a synthesis of limnological knowledge*. Kluwer Academic Publishers, Dordrecht, pp. 551-559.

Northcote, T.G. (2000) Ecological interactions among an Orestiid (Pisces: Cyprinodontidae) species flock in the littoral zone of lake Titicaca. *Advances in ecological research*, 31, 399-420.

Nosil, P., Harmon, L.J. & Seehausen, O. (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, 24, 145-156.

Nyingi, D.W. & Agnèse, J.F. (2007) Recent introgressive hybridization revealed by exclusive mtDNA transfer from *Oreochromis leucostictus* (Trewavas, 1933) to *Oreochromis niloticus* (Linnaeus, 1758) in lake Baringo, Kenya. *Journal of Fish Biology*, 70, 148-154.

Olson, M.E. & Arroyo-Santos, A. (2009) Thinking in continua: beyond the "adaptive radiation" metaphor. *BioEssays*, 31, 1337-1346.

Orlove, B.S., Levieil, D.P. & Treviño, H. (1992) Social and economic aspects of the fisheries. *In:* Dejoux, C. & Iltis, A. (Eds.) *Lake Titicaca: a synthesis of limnological knowledge*. Kluwer Academic Publishers, Dordrecht, pp. 500-504.

Palumbi, S.R. (1994) Genetic divergence, reproductive isolation and marine speciation. *Annual Review of Ecology and Systematics*, 25, 547-572.

Parenti, L.R. (1981) A phylogenetic and biogeographic analysis of cyprinodontiform fishes (Teleostei, Atherinomorpha) *Bulletin of the American Museum of Natural History*, 168, 335-557.

Parenti, L.R. (1984a) Biogeography of the Andean killifish genus *Orestias* with comments on the species flock concept. *In:* Echelle, A.A. & Kornfield, I. (Eds.) *Evolution of fish species flocks.* University of Maine, Orono, pp. 85-92.

Parenti, L.R. (1984b) A taxonomic revision of the Andean killifish genus *Orestias* (Cyprinodontiformes, Cyprinodontidae). *Bulletin of the American Museum of Natural History*, 178, 107-214.

Parker, A. & Kornfield, I. (1995) Molecular perspective on evolution and zoogeography of cyprinodontid killifishes (Teleostei; Atherinomorpha). *Copeia*, 1995, 8-21.

Placzek, C., Quade, J. & Patchett, P.J. (2006) Geochronology and stratigraphy of late Pleistocene lake cycles on the southern Bolivian Altiplano: Implications for causes of tropical climate change. *Geological Society of America Bulletin*, 118, 515-532.

Pollard, D.A., Iyer, V.N., Moses, A.M. & Eisen, M.B. (2006) Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. *PLOS Genetics*, 2, 1634-1647.

Polo Sanchez, M.E. (2005) Reproducción artificial e incunacion artesanal *in situ* del qañu (*Orestias albus*) para fines de recuperacion. *In*. Facultad de Agronomía, Universidad Mayor de San Andrés, Tésis de Licenciatura, La Paz, pp. 151.

Quattro, J.M. & Jones, W.J. (1999) Amplification primers that target locus-specific introns in Actinopterygian fishes. *Copeia*, 1999, 191-196.

Raeymaekers, J.A.M., Raeymaekers, D., Koizumi, I., Geldof, S. & Volckaert, F.A.M. (2009) Guidelines for restoring connectivity around water mills: a population genetic approach to the management of riverine fish. *Journal of Applied Ecology*, 46, 562-571.

Regan, T. (1911) The osteology and classification of the teleostean fishes of the Order Microcyprini. *The Annals and Maganize of Natural History*, 7, 320-327.

Ribbink, A.J. (1984) Is the species flock concept tenable? *In:* Echelle, A.A. & Kornfield, I. (Eds.) *Evolution of fish species flocks*. University of Maine, Orono, pp. 21-25.

Ricciardi, A. & Rasmussen, J.B. (1999) Extinction rates of North American freshwater fauna. *Conservation Biology*, 13, 1220-1222.

Roche, M.A., Bourges, J., Cortes, J. & Mattos, R. (1992) Climatology and hydrology of the lake Titicaca basin. *In:* Dejoux, C. & Iltis, A. (Eds.) *Lake Titicaca: a syntesis of limnological knowledge*. Kluwer Academic Publishers, Dordrecht, pp. 63-88.

Rodriguez-Ramilo, S., Toro, M. & Fernandez, J. (2009) Assessing population genetic structure via the maximisation of genetic distance. *Genetics Selection Evolution*, 41, 49.

Roman, J. & Darling, J.A. (2007) Paradox lost: genetic diversity and the success of aquatic invasions. *Trends in Ecology & Evolution*, 22, 454-464.

Rundell, R.J. & Price, T.D. (2009) Adaptive radiation, nonadaptive radiation, ecological speciation and non ecological speciation. *Trends in Ecology & Evolution*, 24,

Rundle, H.D. & Nosil, P. (2005) Ecological speciation. *Ecology Letters*, 8, 336-352.

Salzburger, W., Baric, S. & Sturmbauer, C. (2002) Speciation via introgressive hybridization in East African cichlids? *Molecular Ecology*, 11, 619-625.

Salzburger, W. & Meyer, A. (2004) The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. *Naturwissenschaften*, 91, 277-290.

Sanderson, M.J. (1998) Reappraising adaptive radiation. *American Journal of Botany*, 85, 1650-1655.

Schluter, D. (2000) *The ecology of adaptive radiation*. Oxford University Press, Oxford, 288 pp.

Schluter, D. (2009) Evidence for ecological speciation and its alternative. *Science*, 323, 737-741.

Schön, I. & Martens, K. (2004) Adaptive, pre-adaptive and non-adaptive components of radiations in ancient lakes: a review. *Organisms, Diversity & Evolution*, 4, 137-156.

Seehausen, O. (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, 19, 198-207.

Seehausen, O. (2006) African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B: Biological Sciences*, 273, 1987-1998.

Strecker, A.L., Olden, J.D., Whittier, J.B. & Paukert, C.P. (2011) Defining conservation priorities for freshwater fishes according to taxonomic, functional, and phylogenetic diversity. *Ecological Applications*, 21, 3002-3013.

Streelman, J.T. & Albertson, R.C. (2006) Evolution of novelty in the cichlid dentition. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 306, 216-226.

Sturmbauer, C., Baric, S., Salzburger, W., Rüber, L. & Verheyen, E. (2001) Lake level fluctuations synchronize genetic divergences of cichlid fishes in African lakes. *Molecular Biology and Evolution*, 18, 144-154.

Sullivan, J.P., Lavoué, S. & Hopkins, C.D. (2002) Discovery and phylogenetic analysis if a riverine species flock of African electric fishes (Mormyridae: Teleostei). *Evolution*, 56, 597-616.

Sylvestre, F., Servant, M., Servant-Vildary, S., Causse, C., Fournier, M. & Ybert, J.-P. (1999) Lake-level chronology on the southern Bolivian Altiplano (18°-23°S) during late-glacial time and the early Holocene. *Quaternary Research*, 51, 54-66.

Takahashi, K., Terai, Y., Nishida, M. & Okada, N. (2001) Phylogenetic relationships and ancient Incomplete lineage sorting among cichlid fishes in lake Tanganyika as revealed by analysis of the insertion of retroposons. *Molecular Biology and Evolution*, 18, 2057-2066.

Takahashi, T. & Koblmmüller, S. (2011) The adaptive radiation of cichlid fish in lake Tanganyika: a morphological perspective. *International Journal of Evolutionary Biology*, 2011,

Takahashi, T. & Koblmüller, S. (2011) The adaptive radiation of cichlid fish in lake Tanganyika: a morphological perspective. *International Journal of Evolutionary Biology*, 2011, 1-14.

Tchernavin, V.V. (1944) A revision of the subfamily Orestiinae. *Proceedings of the Zoological Society of London*, 114, 140-233.

Treviño, H., Torres Caleron, J. & Roncal Gutierrez, M. (1992) The fishery potential. *In:* Dejoux, C. & Iltis, A. (Eds.) *Lake Titicaca: a synthesis of limnological knowledge* Kluwer Academic Publishers, Dordrecht, pp. 539-549.

Väinölä, R., Witt, J.D.S., Grabowski, M., Bradbury, J.H., Jazdzewski, K. & Sket, B. (2008) Global diversity of amphipods (Amphipoda; Crustacea) in freshwater. *Hydrobiologia*, 595, 241-255.

Valenciennes, A. (1839) Poissons d'Amerique. Rapport sur quelques poissons d'Amérique rapportés par M. Pentland du lac Titicaca et des autres points élevés des Andes. *L'Institut, ann.*, 7, 118.

Valenciennes, A (1846) Rapport sur quelques poissons d'Amérique rapportés par M. Pentland. *L'Institut*, 7, 118.

Van Damme, P.A., Carvajal-Vallejos, F., Sarmiento, J., Barrera, S., Osinaga, K. & Miranda-Chumacero, G. (2009) Peces. *In: Libro rojo de la fauna silvestre de vertebrados de Bolivia*. Ministerio de Medio Ambiente y Agua, La Paz, pp. 25-90.

Van Der Voo, R. (1993) *Paleomagnetism of the Atlantic, Tethys, and Iapetus oceans*. Cambridge University Press, Cambridge, 424 pp.

Vila, G. (1975) Geología de los depositos salinos Andinos, provincia de Antofagasta, Chile. *Revista Geologica de Chile*, 2, 41-55.

Vila, I. (2006) A new species of killifish in the genus *Orestias* (Teleostei: Cyprinodontidae) from the Southern High Andes, Chile. *Copeia*, 2006, 472-477.

Vila, I., Mendez, M.A., Scott, S., Morales, P.M. & Poulin, E. (2007a) Threatened fishes of the world: *Orestias ascotanensis* Parenti, 1984 (Cyprinodontidae). *Environmental Biology of Fishes*, 80, 491-492.

Vila, I., Morales, P., Scott, S., Poulin, E., Véliz, D., Harrod, C. & Méndez, M.A. (2013) Phylogenetic and phylogeographic analysis of the genus *Orestias* (Teleostei:

Cyprinodontidae) in the southern Chilean Altiplano: the relevance of ancient and recent divergence processes in speciation. *Journal of Fish Biology*, 82, 927-943.

Vila, I., Pardo, R. & Scott, S. (2007b) Freshwater fishes of the Altiplano. *Aquatic Ecosystem Health & Management*, 10, 201-211.

Vila, I. & Pinto, M. (1986) A new species of Killifish (Pisces, Cyprinodontidae) from the Chilean Altiplano. *Revue d'Hydrobiologie Tropicale*, 19, 233-239.

Vila, I., Scott, S., Iturra, P., Lam, N. & Méndez, M.A. (2010) Karyological and morphological analysis of divergence among species of the killifish genus *Orestias* (Teleostei: Cyprinodontidae) from the southern Altiplano. *In:* Nelson, J.S., Schultze, H.P. & Wilson, M.V.H. (Eds.) *Origin and Phylogenetic Interrelationships of Teleosts*, Munich, pp. 471-480.

Vila, I., Scott, S., Mendez, M.A., Valenzuela, F., Iturra, P. & Poulin, E. (2011) *Orestias gloriae*, a new species of cyprinodontid fish from saltpan spring of the southern high Andes (Teleostei: Cyprinodontidae). *Ichthyological Exploration of Freshwaters*, 22, 345-353.

Villwock, W. (1962) Die Gattung *Orestias* (Pisces, Microcyprini) und die Frage der intralakustrischen Speziation im Titicaca-Seengebiet. *Verhandlungen der Deutschen Zoologischen Gesellschaft Wien*, 54, 610-624.

Villwock, W. (1986) Speciation and adaptive radiation in the Andean *Orestias* fishes. *In:* Vuilleumier, F. & Monasterio, M. (Eds.) *High altitude tropical biogeography*. Oxford University Press, Oxford, pp. 387-403.

Villwock, W. & Sienknecht, U. (1993) Die zahnkarpfen der gattung *Orestias* (Valenciennes, 1839) (Teleostei: Cyprinodontidae) aus dem Altiplano von Bolivien, Chile und Peru. Ein beitrag zur entstehung ihrer formenvielfalt und

intragenerischen verwandtschaftsbeziehungen. *Mitteilungen aus dem Hamburgischen Zoologischen Museum und Institut*, 90, 321-362.

Villwock, W. & Sienknecht, U. (1995) Intraspezifische variabilität im genus *Orestias* Valenciennes, 1839 (Teleostei: Cyprinodontidae) und zum problem der artidentität. *Mitteilungen aus dem Hamburgischen Zoologischen Museum und Institut*, 92, 381-398.

Villwock, W. & Sienknecht, U. (1996) Contribution to the knowledge and history of Chilean fishes. The cyprinodontids of the genus *Orestias* Val. 1839 (Teleostei: Cyprinodontidae) of the Chilean Altiplano. *Medio Ambiente*, 13, 119-126.

Von Rintelen, K., Glaubrecht, M., Schubart, C.D., Wessel, A. & Von Rintelen, T. (2010) Adaptive radiation and ecological diversification of Sulawesis's ancient lake shrimps. *Evolution*, 64, 3287-3299.

Wagner, C.E., Harmon, L.J. & Seehausen, O. (2012) Ecological opportunity and sexual selection together predict adaptive radiation. *Nature*, 487, 366-369.

Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A. & Seehausen, O. (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, 22, 787-798.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R. & Hebert, P.D.N. (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360, 1847-1857.

Wirrmann, D. & Mourguiart, P. (1995) Late Quaternary spatio-temporal limnological variations in the Altiplano of Bolivia and Peru. *Quaternary Research*, 43, 344-354.

Multi-scale evolutionary analysis of a high altitude freshwater species flock: diversification of the *agassizii* complex (*Orestias*, Cyprinodontidae, Teleostei) across the Andean Altiplano.
Species flocks —monophyletic groups of diversified endemic species— represent emblematic case-studies for the study of diversification (e.g. East African Cichlidae). Two main evolutionary processes at the origin of species flocks have been proposed, including adaptive radiation (rapid ecological speciation) and non-adaptive radiation (slower mutation-order speciation). I here study a rare case of high-altitude freshwater species flock endemic to the inter-Andean basin, namely the most diversified species complex of the genus *Orestias*: the *agassizii* complex. At the phylogenetic level, the molecular analyses integrating type specimens (i) allowed the further refinement of the *agassizii* complex, which should exclude the *luteus* group and *O. jussiei*, and (ii) severely questioned the validity of the diagnostic morphological characters traditionally used to delimit complexes and species of *Orestias*. The 'deep and shallow' phylogenetic pattern observed within the genus (two divergent lineages able to hybridize and diversifying into non-monophyletic morpho-species) fits a scenario of ongoing ecological speciation within Lake Titicaca likely starting in the Pleistocene. At the phylogeographic/population genetic level, the results suggest that paleolake fluctuations and connectivity among basins had a major impact on population structure within the *agassizii* complex. I posit that its diversification outside Lake Titicaca originated from the maximum extensions of paleolakes Ballivian to Tauca (0.6 Mya-14.5 kya) and subsequent dispersal into remote areas, following a 'source-sinks' dynamics affected by isolation-by-distance (notably through altitudinal gradients). The complex is temptatively defined as a 'lacustro-riverine species flock' shaped by both adaptive (Lake Titicaca) and non-adaptive (Altiplano) radiations, pending further eco-morphological and behavioral investigations. In terms of conservation, the phylogenetic redefinition of *Orestias* complexes identified *cuvieri* as an extinction-prone clade (*O. cuvieri* extinct) composed of specialized morphotypes. Within the *agassizii* complex, evolutionary significant units were delimited and restricted to fragile high-altitude habitats across north-eastern and southern Altiplano, but the endemic Chilean morpho-species were not recovered. The accurate level of genetic delimitation reached within the *agassizii* complex and for hybrids (*O. agassizii x O. luteus*) should constitute a useful basis to orientate the restocking programs of *Orestias* species.

Keywords: adaptive radiation, *agassizii* complex, Altiplano, evolution, integrative approach, *Orestias*, species flock

---

Analyse multi-échelle d'un essaim d'espèces de haute altitude: diversification du complexe *agassizii* (*Orestias*, Cyprinodontidae, Teleostei) dans l'Altiplano andin.
Les essaims d'espèces —groupes monophylétiques d'espèces endémiques diversifiées— constituent des cas emblématiques pour l'étude de la diversification (e.g. les Cichlidae d'Afrique de l'Est). Deux principaux processus évolutifs à l'origine des essaims d'espèces ont été proposés, incluant la radiation adaptative (spéciation écologique rapide) et la radiation non-adaptative ('mutation-order speciation', plus lente). Nous avons étudié un cas unique d'essaim d'espèces d'eau douce de haute altitude endémique du bassin inter-andin, à savoir le complexe d'espèces le plus diversifié au sein du genre *Orestias* : le complexe *agassizii*. Au niveau phylogénétique, les analyses génétiques intégrant les spécimens types (i) ont permis de délimiter plus avant le complexe *agassizii*, qui devrait exclure le groupe *luteus* et *O. jussiei*, et (ii) remettent en question la validité des caractères morphologiques diagnostiques traditionnellement utilisés pour délimiter les complexes et les espèces d'*Orestias*. Le patron phylogénétique 'deep and shalllow' observé au sein du genre (deux lignées divergentes capables de s'hybrider et se diversifiant en morpho-espèces non monophylétiques) correspond à un scénario de spéciation écologique en cours au sein du Lac Titicaca commençant vraisemblablement au Pléistocène. Au niveau phylogéographique / génétique des populations, les résultats suggèrent que les fluctuations des paléolacs et la connectivité entre les bassins ont eu un impact majeur sur la structure des populations au sein du complexe *agassizii*. Nous émettons l'hypothèse que les extensions maximales des paléolacs Ballivian à Tauca (0.6 Mya-14.5 kya) suivies d'évènements ultérieurs de dispersion vers des zones éloignées sont à l'origine de sa diversification à l'extérieur du Lac Titicaca, suivant une dynamique 'source-puits' influencée par l'isolement par la distance (notamment par les gradients altitudinaux). Le complexe est provisoirement défini comme un 'essaim d'espèces lacustro-riverin' résultant à la fois de radiations adaptative (Lac Titicaca) et non-adaptative (Altiplano), dans l'attente de futures investigations éco-morphologiques et comportementales. En termes de conservation, la redéfinition phylogénétique des complexes d'*Orestias* a permis d'identifier *cuvieri* comme un clade vulnérable à l'extinction (*O. cuvieri* éteint) et composé de morphotypes spécialisés. Au sein du complexe *agassizii*, des 'evolutionary significant units' ont été délimitées dans des habitats fragiles de haute altitude au nord-est et au sud de l'Altiplano, mais les morpho-espèces endémiques chiliennes n'ont pas été confirmées. Le niveau précis de délimitation génétique atteint au sein du complexe *agassizii* et pour les hybrides (*O. agassizii x O. luteus*) devrait constituer un cadre utile à l'orientation des programmes de repeuplements des espèces d'*Orestias*.

Mots clés : radiation adaptive, complexe *agassizii*, Altiplano, évolution, approche intégrative, *Orestias*, essaim d'espèces