GENETIC CONSEQUENCES OF LIVING IN A HIGHLY POLLUTED ENVIRONMENT: THE CASE OF THE SILVERSIDE *BASILICHTHYS MICROLEPIDOTUS* (JENYNS) (TELEOSTEI: ATHERINOPSIDAE) IN THE MAIPO RIVER BASIN, CENTRAL CHILE

Tesis

Entregada a La

Universidad De Chile

En Cumplimiento Parcial De Los Requisitos

Para Optar Al Grado De

Doctor en Ciencias con Mención en Ecología Y Biología Evolutiva

Facultad De Ciencias

Por

Caren Vega Retter

Julio, 2014

Director de Tesis Dr.: David Véliz

Co-director de Tesis: MSc. Irma Vila

FACULTAD DE CIENCIAS

UNIVERSIDAD DE CHILE

INFORME DE APROBACION

TESIS DE DOCTORADO

Se informa a la Escuela de Postgrado de la Facultad de Ciencias que la Tesis de Doctorado presentada por la candidata.

Caren Vega Retter

Ha sido aprobada por la comisión de Evaluación de la tesis como requisito para optar al grado de Doctor en Ciencias con mención en Ecología Y Biología Evolutiva, en el examen de Defensa Privada de Tesis rendido el día 26 de Junio de 2014

Director de Tesis:

Dr. David Véliz

Co-director de Tesis:

MSc. Irma Vila

Comisión de Evaluación de la Tesis

Dr. Elie Poulin .....

Dr. Marco Méndez .....

Dr. Brian Dyer .....

A mis abuelos, mi fortaleza

BIOGRAFÍA

Caren Vega Retter, nació el 21 de Febrero de 1985 en Stralsund, Alemania. Su crianza estuvo fuertemente marcada por la mezcla cultural Chilena-Alemana de sus padres, Gerit Vega y Hernán Vega. En Octubre de 1990, cuando Caren tenía 5 años la familia se traslada a Chile para comenzar una nueva vida. Su enseñanza media la realizó en el colegio Compañía de María Seminario, siempre bajo la fuerte influencia de sus abuelos, Hernán Vega Campos y Viela Riquelme Pinna. Después de terminar el colegio estudio un año la carrera de Biólogo en la Pontificia Universidad Católica de Valparaíso, para posteriormente en el año 2004 comenzar la Carrera de Licenciatura en Ciencias c/m en Biología en la Universidad de Chile, carrera que finaliza el año 2007. En Marzo del 2009 comienza su Doctorado en Ciencias c/m en Ecología y Biología Evolutiva en la Universidad de Chile, financiada por una Beca Doctoral de CONICYT. Caren es una apasionada por la ciencia, pero también por la educación, se ha desempeñado como profesora tanto en colegios como en instituciones de educación superior.

AGRADECIMIENTOS

Quiero agradecer a mi tutor, el Dr. David Véliz, por su constante apoyo y guía durante la realización de esta tesis doctoral, además de permitirme realizar una investigación científica de mi propio interés. También quiero agradecer a mi co-directora de tesis, la profesora Irma Vila, por su constante aliento durante todo el desarrollo de mi tesis. Además, quiero agradecer a todo el equipo de trabajo del Laboratorio de Limnología por sus constantes sugerencias y aportes a mi tesis, y especialmente a Claudio Quezada, Pablo Muñoz, Juan Pablo Oyanedel y Sergio Scott por su ayuda en terreno y a José Luis Carvajal por su apoyo y amistad durante toda esta etapa.

Mis agradecimientos al Dr. Arne Nolte, jefe del laboratorio "Evolutionary genetics of fishes" del Max-Planck Institute for Evolutionary Biology, ya que su ayuda y enseñanzas fueron fundamentales para el desarrollo y finalización de la última parte de mi investigación.

Además, quiero agradecer a mi familia: mis padres, abuelos, tíos y a mi hermano y su familia por el constante apoyo en esta etapa y por incentivarme permanentemente en todo lo que emprendo. Agradezco a mis amigas de la vida: Millaray, Marcela, Tsiu, Denisse y Karen por animarme en este largo camino y por intentar constantemente comprender mi trabajo.

Finalmente, quiero agradecer a Juan Pablo por ser mi soporte, apoyo e inspiración durante todo este proceso y por su apoyo en los momentos difíciles de esta travesía, sin él este camino habría sido mucho más difícil.

Esta tesis fue financiada por Beca Doctoral CONICYT 21090188, Beca CONICYT Apoyo de Tesis Doctoral 24121005, Beca Chile de Pasantía en el extranjero 75130012, Beca CONICYT de término de Tesis Doctoral y Beca MECESUP: Consolidación e internacionalización del programa de doctorado en Ecología y Biología Evolutiva (EBE), UCH 0803.

SUBJECT INDEX

TABLE LIST

Chapter 3

FIGURE LIST

Chapter 1

whiskers are the minimum and maximum values. The band inside the box is the median (second quartile), and the points represent the outliers.

Figure 4 Boxplot of the historical and current effective population size (N_e) estimated for populations of *B. microlepidotus* in the Maipo River basin. The box represents the first and third quartiles, and the whiskers are the minimum and maximum values. The band inside the box is the median (second quartile), and the points represent the outliers.

Chapter 2

Figure 1 Sampling sites of *B. microlepidotus* in the Limarí, Choapa, Aconcagua and Maipo River basins. In the Limarí basin: HU: Huampulla; PA: Paloma; CM: Caballo Muerto; within Choapa basin: CH: Chillepin; CA: Camisas; CP: Choapa Pueblo; in the Aconcagua basin: PO: Pocuro; PN: Panuehue; LL: LlaLlay; in the Maipo basin: PU: Puangue; PEL: Pelvin; PN: Peñaflor; IM: Isla de Maipo; MEL: Melipilla; MA: Maipo; SFM: San Francisco de Mostazal.

Figure 2 Example of an analysis performed with DFDIST implemented in MCHEZA. Distribution of F_{ST} values for each locus as a function of locus heterozygosity for the comparison of the IMPN – MEL (NP-P) population pair. Candidate loci identified by DFDIST are located in the dark gray (positive selection) and gray (balancing selection) regions with neutral loci in the middle region (light gray). Loci indicated by an arrow are

those identified as under directional selection at 95% confidence level and with FDR < 0.1.

Figure 3 Example of an analysis performed with BAYESCAN: plot of F_{ST} values against posterior probability estimates for the comparison of the PU – PEL (NP-P) population pair. The labelled marker (locus 72) is a locus detected as candidate loci under directional with a posterior probability of 0.95 and FDR of 0.05.

Figure 4 a) Venn diagram summarizing the number of loci detected under directional selection by one or both methods used, DFDIST and BAYESCAN; b) Venn diagram summarizing the number of loci detected as under directional selection related to pollution in one or both polluted populations (MEL and PEL).

Figure 5 Discriminant Analysis of Principal Components for a) candidate loci under directional selection in PEL population, b) candidate loci under directional selection in MEL population and c) candidate loci under balancing selection in the PEL and MEL populations.

Chapter 3

Figure 1 Samplig sites of *B. microlepidotus* in the Maipo River basin. Circles represent non-polluted sites and rectangles represent polluted sites according to Vega-Retter et al. (2014).

Figure 2 Principal Component Analysis performed with 15 environmental variables measured in the four sampling sites of *B. microlepidotus*. The non-polluted group contains the following sites: SFM, IM; the polluted group contains the MEL and PEL sites. PC1 and PC2 stand for Principal Components 1 and 2, respectively.

RESUMEN

En la actualidad, la actividad humana está produciendo grandes cambios en la biodiversidad, disminuyendo el tamaño de las poblaciones aumentando la vulnerabilidad y el riesgo de extinción de las especies. Debido a esto, la adaptación de los organismos a las nuevas condiciones ambientales se ha convertido en uno de los mayores focos de investigación en el área de la conservación, y la comprensión de la base genética de esta adaptación es actualmente un tema fundamental en la biología evolutiva. Los sistemas de agua dulce son uno de los ambientes que han sido más impactados por la actividad humana, siendo la contaminación uno de los factores más importantes. En Chile, una de las cuencas más perturbadas y con altos niveles de contaminación es la cuenca del río Maipo, en la cual se encuentra ubicada la cuidad de Santiago. En esta cuenca habita el pejerrey de río endémico *Basilichthys microlepidotus* (Jenyns). Un trabajo previo demuestra que esta especie se comporta como una sola gran población en tres cuencas de la zona norte del país, sin embargo se observó una diferenciación poblacional para esta especie en la cuenca del río Maipo. Considerando esta información, el objetivo principal de esta tesis fue determinar las consecuencias genéticas de la contaminación en las poblaciones de *B. microlepidotus* de la cuenca del río Maipo. Para esto, se determinó la estructuración poblacional del pejerrey con la ayuda de la variabilidad de ocho loci microsatélites. Además, se realizó un escaneo genómico para determinar la proporción de loci bajo selección direccional y balanceada debido a la contaminación, esto con el uso de marcadores AFLP. Finalmente, se identificó genes con expresión diferencial y polimorfismos de nucleótido simple (SNP) relacionados a la contaminación utilizando la

técnica de RNA-seq. En total, cinco poblaciones de pejerrey fueron encontradas en la cuenca del río Maipo, dos habitando zonas contaminadas y tres habitando zonas no contaminadas. No se observó efectos de la contaminación en la diversidad génica o tasa de migración de las poblaciones de *B. microlepidotus*, pero un total de diez loci de AFLP (6.7% del total de loci) fueron identificados como candidatos de estar bajo selección debido a la contaminación; seis (4.0%) bajo selección direccional y cuatro (2.7%) bajo selección balanceada. Además, se detectaron seis genes con expresión diferencial, dos sub-regulados en los individuos de las áreas contaminadas y cuatro sobre-regulados en estos individuos. También se identificó dos SNPs posiblemente relacionados con la contaminación, ambos ubicados en genes mitocondriales. Se concluye que la contaminación está causando un efecto selectivo, adaptativo y fisiológico sobre las poblaciones de *B. microlepidotus* habitando en la cuenca del río Maipo.

ABSTRACT

At the present, human activities produce changes in the biodiversity by reducing the population size that increases vulnerability and extinction risk of species. Thus, the organismal adaptation to new environmental conditions has become a major research focus in conservation and the understanding of the genetic basis of this adaptation is currently a fundamental issue in evolutionary biology. Freshwater systems are one of the environments that has been most impacted by human activity, being the pollution one of the most important factors affecting rivers and lakes. In Chile, one of the basins most perturbed and with high pollution levels is the Maipo River basin where the city of Santiago is located. The silverside *Basilichthys microlepidotus* (Jenyns) is an endemic fish species that inhabits this basin. A previous work performed with this species showed that each river drainage studied in northern of Chile contained one population, however several populations were detected inside of the Maipo River basin. Considering this information, the main goal of this research was to determine the genetic consequences of the pollution on the different populations of the silverside *B. microlepidotus* in the Maipo River basin. For this, the population structure of the silverside was first identified by analyzing the variability of eight microsatellites loci, plus a genome scan with AFLP loci was performed in order to determine the proportion of loci under directional and balancing selection related to pollution. Finally, genes with differential expression and single nucleotide polymorphisms (SNP) related to the pollution were identified by using the RNA-seq technique. In total, five populations of *B. microlepidotus* were detected in the Maipo River basin, two inhabiting in polluted

areas and three in non-polluted areas. No effects of the pollution over the gene diversity or migration rate were observed, however a total of ten ALFP loci (6.7% of the total loci) were identified as candidate loci under selection related with the pollution; six (4.0%) under directional selection and four (2.7%) under balancing selection. Moreover, six genes with differential expression were identified, two down-regulated in the individuals from the polluted areas and four up-regulated in these individuals. Also, two candidate SNPs possibly related with the pollution and located in mitochondrial genes were detected. It is concluded that pollution is having a selective, adaptive and physiological effect over the populations of *B. microlepidotus* inhabiting in the Maipo River basin.

GENERAL INTRODUCTION

There is no doubt that human activities affect biodiversity. Some of the most documented consequences are the presence of invasive species that displace native species, changes in atmospheric composition causing climate change, the increasing nitrogen deposition and habitat fragmentation (Dukes and Mooney 1999; Sala et al. 2000; Fagúndez 2013). Human activity has thus produced modifications in the distribution of populations, vulnerability and extinction of species. Therefore, whereas the organismal adaptation to new environmental conditions has become a major research focus in conservation, the understanding of the genetic basis of this adaptation is currently a fundamental issue in evolutionary biology.

Specifically, freshwater systems are one of the environments that has been impacted the most by human activity, with more than 20% of freshwater fish diversity in a state of threatened or extinct ten years ago (Jackson et al. 2001). The main effects come from overfishing, changes in water flow, invasive species, pollution and habitat fragmentation (Fahrig 1997; Abell et al. 2007; Mora et al. 2007). In this context, pollution is one of the most important factors affecting the freshwater systems becoming the principal source of water quality degradation in the world (World Water Assessment Programe 2009).

Pollution affects the abundance of aquatic biota due to increasing mortality, reduction in fertility and changes in behavior and morphology (Weis et al. 2001; Fracácio et al. 2003;

Pandey et al. 2008). It has also been associated with changes in the effective size, genetic diversity and mutation rate of natural populations (Bickham et al. 2000; Theodorakis 2003). In the case of genetic diversity, evidence has been controversial (Belfiore and Anderson 2001; DiBattista 2008). There is clear evidence that pollution has added new genetic variations by increasing mutation rates (Baker et al. 2001). However, other researchers have also reported that pollution decreases population genetic diversity by reducing population size (Murdoch and Hebert 1994; Bickham et al. 2000; Domínguez-Domínguez et al. 2007). Furthermore, contradictory evidence about the effect of pollution on gene flow has been reported. For example, Theodorakis et al. (2006) found that the gene flow levels of the fish *Lepomis auritas* (Linnaeus) were influenced by pollution. Conversely, Durrant et al. (2011) showed that pollution was not a barrier for *Salmo trutta* (Linnaeus). Thus, available evidence indicates that the effect of pollution on fish populations is not consistent among different studies. Therefore, it is important to study each case individually.

Pollution can also act as a strong selective pressure, for this reason, identifying the proportion of genes and the variants (alleles) implied in the specimens survival in contaminated areas is a relevant issue because this is a first step towards a global understanding of the evolutionary processes involved. A useful approach (especially for non-model species) is the genome scan, which allows identify loci showing abnormal levels of structuring and/or polymorphism, potentially as a result of selection (Luikart et al. 2003; Beaumont 2005; Storz 2005). For example, in the fish *Fundulus heteroclitus* (Linnaeus) Williams & Oleksiak (2008) found evidences that three populations

inhabiting contaminated sites showed up to 6% of the loci under selection or are loci related to the genome area under selection, while in the freshwater fish *Perca flavescens* (Mitchill) Bélanger-Deschênes et al. (2013) determined that 6.25% of the SNPs studied were under directional selection when fish were exposed to heavy metals. Overall, the pollutants can be implicated in the expression changes of functional genes in the individuals exposed, which play an important role in physiological acclimation and evolutionary adaptation to diverse environments. The new techniques of DNA/RNA analysis, mainly the Next Generation Sequencing, have permitted to identify adaptive genetic variation, especially for the non-model species. Among the type of analysis, RNA-seq is a recently developed approach to analyze the transcriptome profiling. Using RNA-seq technique, Garcia et al. (2012) indicate that there is a complex genomic response in the freshwater fish *Fundulus grandis* (Baird y Girard) exposed to oil, with 1,070 down-regulated and 1,251 up-regulated genes. Huang et al. (2012) investigated the transcriptional response of *Oryzias melastigma* (McClelland) embryos after Perfluorooctane sulfonate (a persistent organic pollutant) exposure; they found differentially expressed genes related to neurobehavioral defects, mitochondrial dysfunction and the metabolism of proteins and fats. On the other hand, Olsvik et al. (2013) found that the gene expression profile of the organisms living in the Lake Mjøsa suggest an enrichment in the mechanism associated with drug metabolism and in the oxidative stress. This evidence on gene expression was associated with the industrial activity, because large quantities of persistent organic pollutants have been discharged into Lake Mjøsa during the last century. Pollution not only produces change in gene expression, also SNPs associated with metal contamination has been detected. Belanger-

Deschenes et al. (2013) determined functional polymorphisms in chronically metal-contaminated wild yellow perch (*Perca* flavescens, (Mitchill); they found three SNPs that could be putatively under selection. Moreover, the allelic frequencies of these three SNPs were mostly correlated with the population mean concentration of hepatic cadmium. Then, the authors proposed that Cd represents a selective agent driving an evolutionary change in the population of the yellow perch inhabiting contaminated areas.

Chilean freshwater systems have a reduced number of fish species and 64% of the 44 native species have been considered to fall within the vulnerable category (Vila et al. 2006). Campos et al. (1998) suggested that habitat fragmentation, invasive species and pollution are the main factors affecting fishes. One of most polluted basins in the country is the Maipo River basin, which has a surface area of 15.304 Km^2, extending between parallels 32° 55' - 34° 15' S and 69° 55'-71° 33' W. The principal affluent of this basin is the Maipo River which has 250 Km of length and serves about 70% of the demand for drinking water and about 90% of the irrigation demands (Dirección General de Aguas, 2004), but also this basin is constituted largely by watercourses that drain the water from the snow melt in the Andes Mountains located in the area (Duarte et al., 1971). This basin has around 6.7 million inhabitants according to the most recent census in 2012, which are living in 163 populated areas, including Santiago. Because of this, the basin has experienced water quality deterioration, mainly eutrophication (Pardo et al. 2008), mostly as a product of organic matter from untreated sewage. The first wastewater treatment was implemented in 2001, but it only treats 25% of the polluted

water. Furthermore, this basin contains the largest number of factories in the country, in addition to the presence of mining in the Andes mountains, which adds heavy metals to the Maipo River basin. Consequently, studies on its fauna have shown a significant reduction in richness and abundance of its fish during the last 30 years (Muñoz 2007), suggesting low tolerance of many native fish to pollution. One species that has shown an important decrease in abundance is *Basilichthys microlepidotus* (Jenyns). This species is an atherinopsid endemic to Chile that inhabits lakes and rivers from 28° to 39°S (Véliz et al. 2012) and considered to be endangered (Campos et al. 1998). It is microphagous, feeding on insect larvae, small invertebrates, filamentous algae and detritus (Duarte et al. 1971; Bahamondes et al. 1987); its reproductive period ranges from August to January (Comte and Vila 1992). Previous studies on its population structure performed with the mDNA control region showed that samples from three different sites analyzed within the Limari, Choapa and Aconcagua rivers belong to a single population in each of these basins, suggesting high dispersal capacity. However, genetically different groups were detected in the Maipo River (Quezada-Romegialli et al. 2010).Then, according to this, and taking into account the high perturbation and pollution level of this basin, that areas with different pollution levels were identified at the Maipo River basin (Dirección general de aguas 2012) and that individuals of *B. microlepidotus* have been observed throughout almost all the basin, the objectives of this research were: i) to identify the population structure of *B. microlepidotus* within the basin, ii) to assess the effect of the pollution on both the gene diversity and the migration rate of the *B. microlepidotus* within the basin, iii) to determine the proportion of loci under selection in *B. microlepidotus* related with the pollution within the Maipo River basin and iv) to assess

the difference in a transcriptomical expression and single polymorphism nucleotide among individuals of *B. microlepidotus* inhabiting non-polluted and polluted areas.

Hypothesis

Considering that areas with high pollution and low pollution were identified in the Maipo River basin and that the silverside *Basilichthys microlepidotus* inhabit most of this basin, the general hypothesis of this study is: Pollution has genetic, selective and physiological effects on the populations of *B. microlepidotus* inhabiting the Maipo River basin.

Followed by the three specific hypotheses:

Hypothesis 1: Pollution has affected the gene diversity and migration rates of the populations of *B. microlepidotus* inhabiting the Maipo River basin. (Chapter 1)

Hypothesis 2: Pollution acts as a selective factor changing the allelic composition of populations of *B. microlepidotus* inhabiting polluted sites compared with populations located in non-polluted areas within the Maipo River basin. (Chapter 2)

Hypothesis 3: Pollution affects the gene expression of the *B. microlepidotus* inhabiting the Maipo River basin. (Chapter 3)

General objective

- Determine the genetic effects of the pollution on populations of the silverside *Basilichthys microlepidotus* in the Maipo River basin.

Specific objectives

- Characterize the pollution level of the sites where populations of the *B. microlepidotus* inhabit in the Maipo River basin. (Chapter 1)

- Identify the population structure of *B. microlepidotus* within the Maipo River basin. (Chapter 1)

- Determine differences in gene diversity and gene flow in populations of the silverside inhabiting polluted sites compared to this in non-polluted areas. (Chapter 1)

- Determine the proportion of loci that are under or are related to areas of the genome under directional selection in populations of *B. microlepidotus* located in polluted and non-polluted areas. (Chapter 2)

- Determine the proportion of loci under o related to areas of the genome under balancing selection in *B. microlepidotus* located in polluted areas. (Chapter 2)

- Identify genes with differential expression between individuals of *B. microlepidotus* inhabiting polluted and non-polluted sites. (Chapter 3)

- Identify Single Nucleotide Polymorphism (SNP) between individuals of *B. microlepidotus* inhabiting polluted versus non-polluted sites. (Chapter 3)

REFERENCES

Abell R, Allan JD, Lehner B (2007) Unlocking the potential of protected areas for freshwaters. Biol Conserv 134: 48-63.

Bahamondes I, Soto D, Vila I (1979) Hábitos alimentarios de las especies de Atherinidae del Embalse Rapel. Medio Ambiente 4: 3-18.

Baker RJ, Bickham AM, Bondarkoy M, Gascjak SP, Matson CW, Rodgers BE, Wickcliffe JK, Chesser RK (2001) Consequences of polluted environments on population structure: the bank vole (*Clethrionomys glareolus*) at Chernobyl. Ecotoxicol 10: 211-216.

Beaumont MA (2005) Adaptation and speciation: what can FST tell us? Trends in Ecology & Evolution 20: 435–440.

Bélanger-Deschênes S, Couture P, Campbell PGC, Bernatchez L (2013) Evolutionary change driven by metal exposure as revealed by coding SNP genome scan in wild yellow perch (*Perca flavescens*). Ecotoxicology 22: 938-957.

Belfiore NM, Anderson SL (2001) Effects of contaminants on genetic patterns in aquatic organism: a review. Mutat Res 489: 97-122.

Bickham JW, Sandhu S, Hebert PDN, Chikhi L, Athwal R (2000) Effects of chemical contaminants on genetic diversity in natural populations: implications for biomonitoring and ecotoxicology. Mutat Res 463: 33-51.

Campos H, Dazarola G, Dyer B, Fuentes L, Gavilan J, Huaquin L, Martinez G, Melendez R, Pequeño G, Ponce F, Ruiz V, Siefeld W, Soto D, Vega R, Vila I (1998) Categorías de conservación de peces nativos de aguas continentales de Chile. Boletín del Museo Nacional de Historia Natural 47: 101-122.

Comte S, Vila I (1992) Spawning of *Basilichthys microlepidotus* (Jenyns). Journal of Fish Biology 41: 971-981.

DiBattista JD (2008) Patterns of genetic variation in anthropogenically impacted populations. Conserv Genet 9: 141-156.

Dirección General de Aguas (2004) Diagnostico y clasificación de los cursos y cuerpos de agua según objetivos de calidad, cuenca del Rio Maipo. Ministerio de Obras Públicas, Gobierno de Chile.

Dirección General de Aguas (2012) Desarrollo metodológico e informe con antecedentes para la realización de los análisis de impacto económico y social de las normas secundarias de calidad de aguas de las cuencas Cachapoal, Maipo y Elqui. Volumen n° 3: Cuenca del río Maipo. Ministerio de Obras Públicas, Gobierno de Chile.

Domínguez-Domínguez O, Boto L, Alda F, Pérez-Ponce de León G, Doadrio I (2007) Humans impacts on drainages of the mesa central Mexico, and its genetics effects on an endangered fish, *Zoogoneticus quitzeoensis.* Conserv Biol 21: 168–180.

Duarte W, Feito R, Jara C, Moreno C, Orellana AE (1971) Ictiofauna del sistema hidrográfico del río Maipo. Boletín del Museo Nacional de Historia Natural Chile 32: 227-268.

Dukes JS, Mooney HA (1999) Does global change increase the success of biological invaders? TREE 14: 135-139.

Durrant CJ, Stevens JR, Hogstrand C, Bury NR (2011) The effect of metal pollution on the population genetic structure of brown trout (*Salmo trutta* L.) residing in the River Hayle, Cornwall, UK. Environ Pol. 159: 3595-3603.

Fagúndez J (2013) Heathlands confronting global change: drivers of biodiversity loss from past to future scenarios. Ann Bot-London 111: 151-172.

Fahrig L (1997) Relative effects of habitat loss and fragmentation on population extinction. J Wildl Manag 61: 603-610.

Fracácio R, Verani NF, Gaeta Espíndola EL, Rocha O, Rigolin-Sá O., Andrade CA (2003) Alterations on Growth and Gill Morphology of *Danio rerio* (Pisces, Ciprinidae) Exposed to the Toxic Sediments. Braz Arch of Biol and Technol 46: 685-695.

Garcia T, Shen Y, Crawford D, Oleksiak MF, Whitehead A, Walter RB (2012) RNA-Seq reveals complex genetic response to deepwater horizon oil release in *Fundulus grandis*. BMC Genomics 13: 474.

Huang Q, Dong S, Fang C, Wu X, Ye T, Lin Y (2012) Deep sequencing-based transcriptome profiling analysis of Oryzias melastigma exposed to PFOS. Aquatic Toxicology 120-121: 54-58.

Jackson RB, Carpenter SR, Dahm CN, McKnight DM, Naiman RJ, Postel SL, Running SW (2001) Water in changing world. Issues in Ecol 9: 1-16.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. Nature Reviews Genetics 4: 981–994.

Mora C, Metzger R, Rollo A, Myers RA (2007) Experimental simulations about the effects of overexploitation and habitat fragmentation on population facing environmental warming. Proc of the R Soc of Lond ser B 274: 1023-1028.

Muñoz BM (2007) Cambios en la riqueza íctica del rio Maipo. Tesis Biología con Mención en Medio Ambiente, Facultad de Ciencias, Universidad de Chile.

Murdoch MH, Hebert PDN (1994) Mitochondrial DNA diversity of brown bullhead from contaminated and relatively pristine sites in the Great Lakes. Environ Toxicol and Chem 13: 1281–1289.

Olsvik PA, Berg V, Lyche JL (2013) Transcriptional profiling in burbot(Lota lota) from Lake Mjøsa—A Norwegian Lake contaminated by several organic pollutants. Ecotoxicology and Environmental Safety 92: 94-103.

Pandey S, Parvez S, Ansari RA, et al. (2008) Effects of exposure to multiple trace metals on biochemical, histological and ultrastructural features of gills of a freshwater fish, *Channa punctata* Bloch. Chem-Biol Interact 174: 183–192.

Pardo R, Vila I, Bustamante R (2008) Demographic response of *Stratiodrilus aeglaphilus* (Anelida, Histriobdellidae) to organic enrichment: experimental assessment. Hydrobiologia 598: 11-16.

Quezada-Romegialli C, Fuentes M, Veliz D (2010) Comparative population genetics of *Basilichthys microlepidotus* (Atheriniformes: Atherinopsidae) and *Trichomycterus areolatus* (Siluriformes: Trichomycteridae) in north central Chile. Environ Biol Fish 89:173–186.

Sala OE, Chapin III FS, Armesto JJ, Berlw E, Bloomfield J, Dirzo R, Huber-Sanwald E, Huenneke LF, Jackson RB, Kinzing A, Leemans R, Lodge DM, Mooney HA, Oesterheld M, Poff NL, Sykes MT, Walker BH, Walker M, Wall DH (2000) Global biodiversity scenarios for the year 2100. Science 287: 1770-1774.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. Molecular Ecology 14: 671–688.

Theodorakis CW (2003) Establishing causality between population genetic alterations and environmental contamination in aquatic organisms. Hum Ecol Risk Assess 9: 37-58.

Theodorankis CW, Lee KL, Adams M, Law B (2006) Evidence of altered gene flow, mutation rate, and genetic diversity in redbreast sunfish from a pulp-mill-contaminated river. Environ Sci Technol 40: 377-386.

Véliz D, Catalán L, Pardo R, Acuña P, Diaz A, Poulin E, Vila I (2012) The genus Basilichthys (Teleostei : Atherinopsidae) revisited along its Chilean distribution range (21° to 40° S) using variation in morphology and mtDNA. Revista Chilena de Historia Natural 85: 49-59.

Weis JS, Smith G, Zhou T, Santiago-Bass C, Weis P (2001) Effects of Contaminants on Behavior: Biochemical Mechanisms and Ecological Consequences. Bioscience 51: 209-217.

Williams LM & Oleksiak MF (2008) Signatures of selection in natural populations adapted to chronic pollution. BMC Evolutionary Biology 8: 282-293.

World Water Assessment Programme (2009) The United Nations World Water Development Report 3: Water in a Changing World. UNESCO, Paris and Earthscan, London.

CHAPTER 1

GENETIC EFFECTS OF LIVING IN A HIGHLY POLLUTED ENVIRONMENT: THE CASE OF THE SILVERSIDE *BASILICHTHYS MICROLEPIDOTUS* (JENYNS) (TELEOSTEI: ATHERINOPSIDAE) IN THE MAIPO RIVER BASIN, CENTRAL CHILE

Population Ecology: DOI 10.1007/s10144-014-0444-3.

Caren Vega-Retter[1,2], Pablo Muñoz-Rojas[1,2], Irma Vila[1], Sylvia Copaja[3] & David Véliz[1,2]

[1]Departamento de Ciencias Ecológicas, [2]Instituto de Ecología y Biodiversidad. [3]Departamento de Química, Universidad de Chile, Las Palmeras 3425, Ñuñoa, Santiago, Chile.

ABSTRACT

Freshwater systems are one of the environments most impacted by human activity, with pollution being a highly important factor. In Chile, several rivers exhibit varied levels of pollution, one of which is the Maipo River basin where the city of Santiago is located. The silverside *Basilichthys microlepidotus* (Jenyns) is an endemic fish species that inhabits this basin, thus we hypothesized that pollution has affected gene diversity and migration in populations of *B. microlepidotus* from the Maipo River basin. The aim of this study was to identify the population structure of *B. microlepidotus* in this basin and to determine if the populations of the silverside inhabiting polluted sites present differences in gene diversity and gene flow compared to populations inhabiting non-polluted areas. Using the variability of eight microsatellites, five populations of silverside were detected; three inhabiting non-polluted sites and two inhabiting polluted sites. From this, it was inferred that *B. microlepidotus* has been able to tolerate pollution in the Maipo River basin. No differences in gene diversity or migration were detected between polluted and non-polluted sites but comparison with historical estimation revealed an increase in the current migration rate when all the data from the basin were compared. A reduction in current effective population size was also observed when compared to historical values, and this is probably due to river degradation. Despite the disappearance of other fish species recorded at this basin, our results suggest that *B. microepidotus* is tolerant to pollution, thus indicating that native species respond differently to this environmental factor.

Keywords: Effective population size · Genetic structure · Microsatellites · Migration · Pollution

INTRODUCTION

There is no doubt that human activity affects biodiversity. Some of the most documented consequences of this activity include the introduction of invasive species that displace native species, climate change due to alterations in atmospheric composition, increased nitrogen deposition, and habitat fragmentation (Dukes and Mooney 1999; Sala et al. 2000; Fagúndez 2013). In a wider perspective, the consequences of human activity have modified the distribution and vulnerability of populations and have led to the extinction of numerous species. Therefore, the adaptation of organisms to new environmental conditions has become a major research focus in conservation.

Freshwater systems are one of the environments that has been most impacted by human activity, with more than 20 % of freshwater fish in a state of threatened or extinct (Jackson et al. 2001). This threat to freshwater diversity is mainly the result of overfishing, changes in water flow, invasive species, pollution, and habitat fragmentation (Fahrig 1997; Abell et al. 2007; Mora et al. 2007). Of these, pollution is one of the most important factors affecting freshwater systems because it is the principal source of water quality degradation in the world (World Water Assessment Programme 2009).

Pollution affects the abundance of aquatic biota by increasing mortality, reducing fertility, and changing behaviors and morphology (Weis et al. 2001; Fracácio et al. 2003; Pandey et al. 2008). It has also been associated with changes in the effective size,

genetic diversity, and mutation rate of natural populations (Bickham et al. 2000; Theodorakis 2003). In the case of genetic diversity, evidence has been controversial (Belfiore and Anderson 2001; DiBattista 2008), there is clear evidence that pollution has promoted new genetic variations by increasing mutation rates (Baker et al. 2001), but other researchers have also reported that pollution decreases the genetic diversity of a population by reducing its size (Murdoch and Hebert 1994; Bickham et al. 2000; Domínguez-Domínguez et al. 2007). These differences suggest that species react differently to pollution, although it is important to consider that changes in genetic diversity affect the ability of populations to adapt to future environmental challenges (van Straalen and Timmermans 2002).

Furthermore, contradictory evidence about the effect of pollution on gene flow has been reported. Theodorakis et al. (2006) found that the gene flow levels of the fish *Lepomis auritas* (Linnaeus) were influenced by pollution. Conversely, Durrant et al. (2011) showed that pollution was not a barrier for *Salmo trutta* (Linnaeus). Again, available evidence indicates that the effect of pollution on fish populations is not consistent between different studies or species. Therefore, it is important to research each case individually.

Chilean freshwater systems have a reduced number of fish species, and 64 % of the 44 native species are classified as vulnerable (Vila et al. 2006). Campos et al. (1998) suggested that habitat fragmentation, invasive species, and pollution are the main factors affecting fish in Chile. One of most polluted basins in Chile is the Maipo River basin,

where approximately 40 % of the Chilean population lives, amounting to 6.7 million individuals according to the most recent census in 2012. Because of this, the basin has experienced water quality deterioration, mainly eutrophication (Pardo et al. 2008) as a product of organic matter from untreated sewage. The first wastewater treatment plant for the basin was implemented in 2001, but it only treats 25 % of the polluted water. Furthermore, the Maipo River basin contains the largest number of factories in the country, and mining operations in the Andes Mountains introduce heavy metals into the Maipo River basin. The anthropic disturbances along the Maipo River basin have different components, all with spatial and temporal variations. The principal perturbations include diffuse pollution by sewage, agricultural crops, industrial discharges, pesticides, mining, and gravel extraction (Dirección General de Aguas 2004).

Studies concerning the fauna of the basin have shown a significant reduction in richness and abundance of fish during the last 30 years (Muñoz 2007), suggesting a low tolerance to pollution in many native fish. One species that has shown a marked decrease in abundance is *Basilichthys microlepidotus* (Jenyns). Duarte et al. (1971) found 198 individuals, while Muñoz (2007) collected only 11 specimens at the same 19 sampling sites along the Maipo River basin. This species is an atherinopsid endemic to Chile that inhabits lakes and rivers from 28° to 39°S (Véliz et al. 2012), and it is considered endangered (Campos et al. 1998). This species is microphagous, feeding on insect larvae, small invertebrates, filamentous algae, and detritus (Duarte et al. 1971;

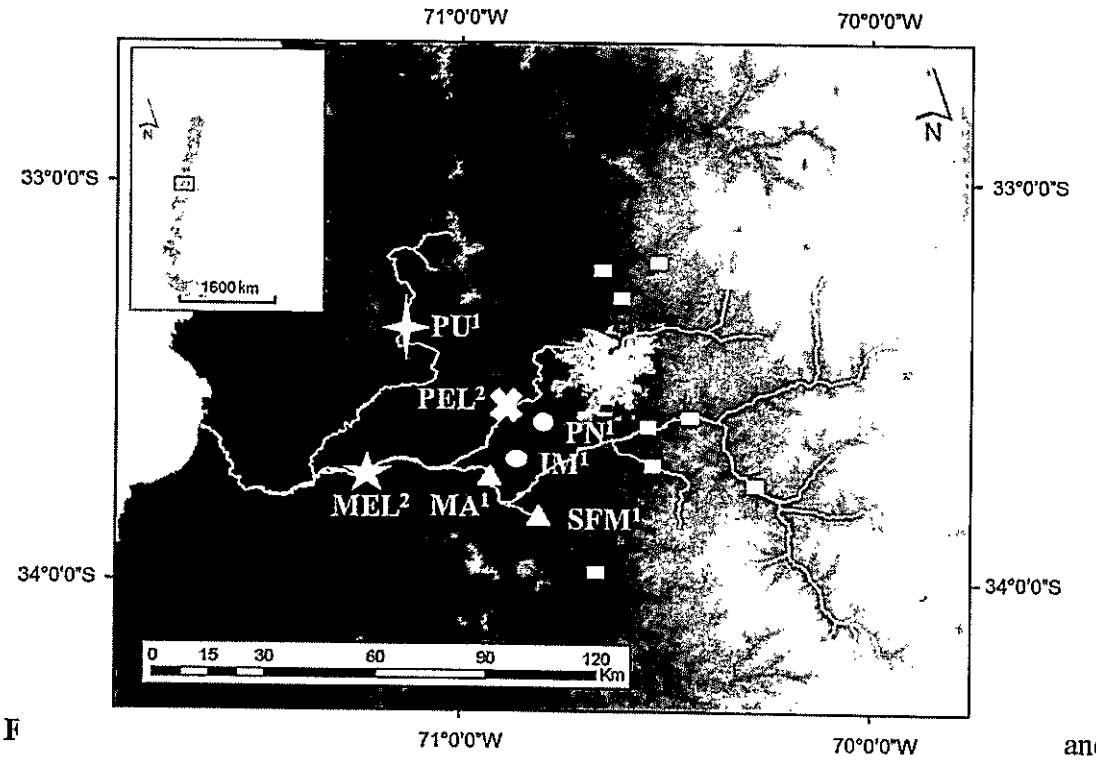
Bahamondes et al. 1987), and its reproductive period ranges from August to January (Comte and Vila 1992).

A previous study performed with the mDNA control region to analyze the population structure of this species showed that samples from the Limari, Choapa, and Aconcagua rivers belonged to a single population within each of these basins, suggesting high dispersal capacity. However, genetically different groups were detected in the Maipo River basin (Quezada-Romegialli et al. 2010). Taking into account all of the particular features of the Maipo River basin, we hypothesized that pollution has affected the gene diversity and migration of populations of *B. microlepidotus* located in the basin. Thus, the goals of this study were i) to characterize the physicochemical water quality of the Maipo River basin sites inhabited by *B. microlepidotus*; ii) to identify the population structure of *B. microlepidotus*; iii) to assess the presence of *B. microlepidotus* at sites with high and low levels of pollution; and iv) to determine the effect of pollution on gene diversity and gene flow of this Chilean silverside in the Maipo River basin.

MATERIALS AND METHODS

Sampling sites and tissue collection

In order to determine the distribution of the silverside *B. microlepidotus* within the Maipo River basin, 16 sites were visited between April 2011 and January 2012. The presence of *B. microlepidotus* was detected at seven of these sites, namely at San Francisco de Mostazal (SFM), Maipo (MA), Puangue (PU), Peñaflor (PN), Pelvin (PEL), Melipilla (MEL), and Isla de Maipo (IM) (Fig. 1). Twenty-four individuals were collected from each site except for MEL, where 22 individuals were collected. Thus, a total of 166 individuals were collected by using a low impact electrofishing device that did not harm the fish; specimens were anesthetized with a 12 mg/L dose of M22 according to Wasko et al. (2003). Tissue samples were obtained by dissecting a small piece of the caudal fin, which was subsequently stored in 99 % ethanol (Merck). To reduce the effect of the anesthesia, fish were maintained for 20 minutes in clean, oxygenated water before being released back into the river.



F and non-polluted sites in the Maipo River basin. Sampling sites: squares are sites that were visited but no fish were observed and, therefore, fish were not collected. The other symbols indicate sampling sites containing the silverside *B. microlepidotus*, where PEL: Pelvin, PU: Puangue, IM: Isla de Maipo, MEL: Melipilla, MA: Maipo, SFM: San Francisco de Mostazal, and PN: Peñaflor. Population structure of *B. microlepidotus*: sampling sites with different symbols represent different populations. Non-polluted and polluted sites: superscript 1 indicates non-polluted sites, superscript 2 indicates polluted sites. The city of Santiago is indicated in grey.

Physicochemical characterization of sampling sites

To determine habitat quality, 19 variables were measured at each site where *B. microlepidotus* was found. *In situ* electrical conductivity (EC), pH, and total dissolved solids (TDS) were measured three times for each site. All measurements were performed with a multiparameter device (Hanna Instruments). Sulfate (SO^{2-}_4), nitrite (NO^-_2), chloride (Cl^-), ammonium (NH^+_4), nitrate (NO^-_3), phosphate (PO^{3-}_4), sodium (Na^+), potassium (K^+), calcium (Ca^{2+}), and magnesium (Mg^{2+}) concentrations were measured in three water samples (1 L) from each site. Each sample was filtered using a cellulose nitrate filter with a 0.45 µm pore diameter (Sartorius) according to Clesceri et al. (2005).

For measurements of the metals copper (Cu), lead (Pb), aluminum (Al), and molybdenum (Mo), a 1 L water sample was collected at each site. All samples were fixed with HNO^-_3 (2 %, Suprapur®Merck) and filtered (0.45 µm diameter) for later analysis by using an atomic absorption spectrophotometer (Shimadzu Spectrophotometer 6800, ASC-6100 autosampler and GFA-EX graphite furnace). To determine dissolved oxygen (DO) concentration, three water samples were collected at each site in 200 mL Nalgene bottles, fixed with manganese sulfate and alkaline iodide, and measured using the Winkler method following the protocol proposed by Strickland and Parsons (1968). Biological Oxygen Demand (BOD_5) was measured to determine the oxygen used by microorganisms to decompose organic residue. This measurement involved collecting 300 mL of water from each site and incubating it at 20 °C for five days before oxygen quantification (Clesceri et al. 2005).

The British Columbia Water Quality Index (BCWQI 1996) was used to determine the degree of pollution at each site. This index is based on the attainment of water quality

objectives, taking into consideration the natural water quality expected at any given site. The index ranks water quality into the following five categories: excellent, good, fair, borderline, and poor. Excellent means that conditions are very close to natural or pristine levels; good indicates that conditions rarely depart from natural or desirable levels; and fair, borderline, or poor mean that conditions sometimes, often, or usually depart from natural or desirable levels, respectively. In order to assign each site to a category, three factors were calculated (F_1, F_2, and F_3). F_1 measure the number of parameters that do not meet the objectives defined according to the expected natural water quality, whereas F_2 is the number of times that the parameters do not attain the objectives. Finally, F_3 is a measure of the maximum quantity by which objectives are not being met. The index was calculated by using the following formula: BCWQI = $[(F_1)^2 + (F_2)^2 + (F_3/3)^2]^{1/2}$. The index was estimated considering all the parameters measured, with exception of those for which information about the objective criteria was not available, which were TDS, NO^-_3, PO^{3-}_4, Na^+, K^+, Ca^{2+}, and Mg^{2+}.

The variables used in the BCWQI are considered significant measurements for the water quality of the Maipo River basin (Dirección General de Aguas, 2004). This physicochemical categorization was complemented with a Principal Component Analysis (PCA), where all the variables measured were included so as to detect the environmental variables that most contributed to the total variance and the location of the sites in each principal component axis. Factor loadings were used to determine the contribution of the variable in each component. This analysis was performed using the ADE4 library implemented in the R software (R Core Team 2013). Finally, for

simplicity of analysis, the sites were classified as polluted or non-polluted according to the results from the BCWQI and PCA analyses.

Microsatellite amplification, descriptive statistics, and population structure

DNA extraction was performed by using the salt extraction method (Aljanabi and Martinez 1997). Four microsatellites described for *Odontesthes perugiae* (Evermann and Kendall), Odon02, Odon07, Odon09, and Odon39 (Beheregaray and Sunnucks 2000), and four microsatellites described for *Odontesthes bonariensis* (Valenciennes), Odon01, Odon19, Odon59, and Odon71 (Koshimizu et al. 2009), were used for population genetic analyses. The PCR reaction mixture had a final volume of 10 µL that included 1.3 µL of 10X PCR buffer (Invitrogen), 0.5 µL of $MgCl_2$ (50 mM) (Invitrogen), 0.5 µL of forward and reverse primers (50 ng/µL) (Applied Biosystems), 2.4 µL of dNTPs (2.5 mM) (Invitrogen), 4.68 µL of H_2O, and 0.12 µL of Taq polymerase (Invitrogen). To this mixture, 1.5 µL of DNA (50 ng/µL) was added.

The PCR touchdown method described by Beheregaray and Sunnucks (2000) was used with the modifications included by Muñoz et al. (2011) for the amplification protocol of the four first microsatellites. Amplification of the loci Odon01, Odon19, Odon59, and Odon71 was performed following the protocol described in Koshimizu et al. (2009). PCR products were genotyped by Macrogen Inc (www.macrogen.com), and the GENEMARKER software was used to build the matrix with allelic data. The

MICROCHECKER software (Van Oosterhout et al. 2004) was used to identify possible genotyping mistakes and the presence of null alleles in the microsatellite data.

To ensure that the individuals were a random subset of the population, relatedness was estimated with the r_{xy} estimator, as indicated by Queller and Goodnight (1989), implemented into the IDENTIX software (Belkhir et al. 2002). Using multilocus genotypic data, IDENTIX estimates relatedness between pairs of individuals within each population. The null hypothesis for the random distribution of related individuals in a sample was also tested by using 1000 permutations of alleles within the respective sample. Linkage disequilibrium was estimated for all pairs of loci, and deviations from the Hardy-Weinberg Equilibrium (HWE) were estimated using the GENETIX software (Belkhir et al. 2000).

Three software programs were used to determine population structure. First, the FLOCK software version 3.0 (Duchesne and Turgeon 2012) was used to determine the most likely number of populations (k) and their membership. The FLOCK software implements a method of repeated re-allocation of collected specimens in order to detect a reliable number of k populations, with each re-allocation being more effective than the previous one in attracting genetically similar individuals. This analysis was run using the random initial partition mode, 20 re-allocations per run, 50 runs, and a minimal log-likelihood (LLOD) of 0.3. In addition, the GENELAND software (Guillot et al. 2005) was used to determine the number of genetically distinct population clusters and their

membership with another criterion. GENELAND uses a Bayesian clustering algorithm to assign individuals to clusters or populations without *a priori* knowledge of the population sampled. This analysis was run 20 times using correlated allelic frequencies with no inferred spatial model. The number of iterations was set to 500000, with a thinning of 500. Finally, the GENETIX software was used to estimate genetic differences between pairs of sites using F_{ST}. GENETIX estimates the F_{ST} based on the multilocus allelic frequency gathered from the individuals collected at each site (Weir and Cockerham 1984). Individual permutations among sites (10000 permutations) were used to estimate the statistical significance of this index. To reduce Type I error, α = 0.01 was used in these paired F_{ST} comparisons. In order to test for a geographical association with population differentiation (F_{ST}), a Mantel test with 10000 permutations was conducted using the GENETIX software. Geographical distances between pairs of sites were estimated using Google Earth, following the river line.

After that the spatial structure of populations was estimated, it was associated with the environmental condition (polluted or non-polluted) as determined on the basis of BCWQI and PCA results. The number of alleles (N_A), allelic richness (AR), and gene diversity (GD) were estimated for each of the populations detected. The N_A was estimated by using the GENETIX software, while AR and GD were estimated using the FSTAT software (Goudet 1995), which is based on randomization methods to estimate statistical significance of both parameters among populations. In brief, AR is a measure of the number of alleles independent of sample size, while GD is directly calculated from the estimated frequency of alleles per sample. Significant differences in these

parameters were estimated among populations from polluted and non-polluted areas using a permutation ANOVA from the lmPerm library in the R software (R Core Team 2013). Finally, the BOTTLENECK software (Cornuet and Luikart 1996) was used to determine if the detected populations of the silverside had undergone a recent bottleneck. In brief, the BOTTLENECK software estimates the likelihood of recent reductions in effective population size (N_e) by comparing the expected heterocigosity under HWE with the expected heterocigosity under mutation-drift equilibrium. The two-phase mutation model with a 70 % stepwise mutation was used, and the significance was assessed with the Wilcoxon test implemented in the same software.

Historical and contemporary gene flow and effective population size estimates

In order to determine the migration patterns of silverside populations, both historical and contemporary migration rates were estimated. Historical migration rates were first estimated with the MIGRATE software (Beerli and Felsenstein 1999), which uses a coalescent approach to estimate mutation-scaled migration rates (M) for each population analyzed over the last $4N_e$ generations (approximately). The settings used a maximum likelihood, the Brownian motion mutation model, and the matrix migration model containing 20 short chains of 40000 steps and six long chains of 400000 steps, after a burn-in step of 40000. All results were obtained from the average of three independent runs, and all runs were performed at https://www.bioportal.uio.no/. To estimate contemporary gene flow, a Bayesian analysis was performed using the BAYESASS

software (Wilson and Rannala 2003). The BAYESASS software implemented a Bayesian inference, which estimates symmetric or asymmetric migration over the last two to three generations between populations. This analysis was run using 30000000 iterations and a burn-in of 3000000 iterations, with sampling done per 100 iterations. The mixing parameters for allele frequencies, migration rates, and inbreeding coefficients were defined as 0.5, 0.1, and 0.4, respectively. Ten independent runs that started with different seeds were performed to examine consistency in the results. The results were expressed as the average value of these ten independent runs. To make this comparison possible, the historical migration rate was scaled by using the formula $m = M \times \mu$, and the mutation rate used was 2×10^{-3}. This mutation rate is an average value taken from *Lepomis marginatus* (Holbrook) (Mackiewicz et al. 2002) and *Syngnathus typhle* (Linnaeus) (Jones et al. 1999). Both species are phylogenetically close to *B. microlepidotus*.

The historical and recent number of effective migrants per generation (N_em) was estimated using the MIGRATE and BAYESASS programs. The historical N_em was estimated using $M \times \theta/x$ formula, where $\theta = x\ N_e\ \mu$, x depends on the ploidy and inheritance of the data, and for nuclear data is defined as 4. The current N_em was obtained by multiplying the migration rate by the current N_e. The current N_e was calculated as indicated below.

To determine general migration changes within the basin, the difference between the historical and current migration rate for all populations in the Maipo River basin was

tested. In a second analysis, migration comparisons between polluted and non-polluted populations were performed. For this purpose, the immigration/emigration ratio (I/E ratio) was estimated for each polluted and non-polluted population, and the comparisons were performed in the following two ways: i) by assessing statistical differences in the current I/E ratios between non-polluted and polluted populations, and ii) by assessing differences between historical and current I/E ratios for non-polluted and polluted populations. All comparisons were performed for m and N_em using the permutation ANOVA from the lmPerm library in the R software.

The historical effective population sizes of silverside populations were estimated from the θ average, which was estimated from three independent runs performed in the MIGRATE software, for which the settings were a mutation rate = 2×10^{-3} and $x = 4$. An estimation of contemporary effective population sizes was performed through Bayesian approximate computation implemented in ONeSAMP 1.2 (Tallmon et al. 2008) using a minimum and maximum effective population size of 2 and 1000, respectively. To determine differences between the historical and current effective population sizes, a Wilcoxon test was run in the R software.

RESULTS

Physicochemical characterization of sampling sites

The 10 physicochemical features (EC, pH, SO^{2-}_4, NO^-_2, Cl^-, NH^+_4, DO, BOD_5, Cu, and Mo) used in the BCWQI indicated that the IM site was classified as excellent, while four sites (SFM, MA, PU, PN) were ranked as good, and two (PEL and MEL) were classified as fair. For the PCA analysis, Pb and Al were under the limit of detection, and, given this, were not used in the analysis. The PCA analysis was performed with the following seventeen parameters: EC, pH, TDS, SO^{2-}_4, NO^-_2, Cl^-, NH^+_4, NO^-_3, PO_4, Na^+, K^+, Ca^{2+}, Mg^{2+}, Cu, Mo, DO, and BOD_5 (Annex 1). This analysis showed that the first two components accounted for 71 % of variance (Fig. 2), where the Principal Component 1, accounting for 51 % of the total variance, had high loadings for TDS, EC, Cl^-, and SO^{2-}_4. In turn, the Principal Component 2, accounting for 20 % of the variance, had high loadings for Mo, NH^+_4, Cu, and NO^-_2. It is important to note that seven out of the eight variables accounting for most of the variance in the PCA analysis were included in the BCWQI estimation. In addition, the PCA analysis showed a segregation of sites classified as excellent and good from the sites classified as fair (Fig. 2). Taking into account these results and the meaning of each category from the index, the sites with an excellent and good category were defined as non-polluted sites while the sites that were ranked as fair were considered polluted sites (Table 1, Fig. 1).

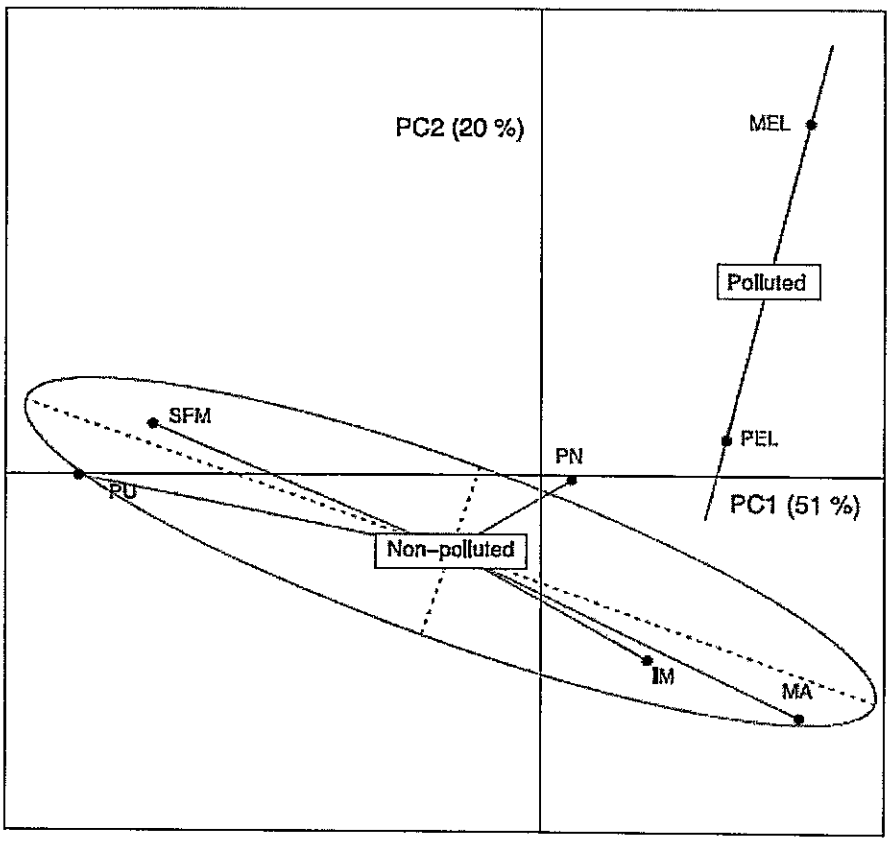


Figure 2. Principal Component Analysis performed with 17 environmental variables measured at the seven sampling sites where *B. microlepidotus* was present. The ellipse from the non-polluted group contains the SFM, MA, PU, IM, and PN sites; whereas the polluted group contains the MEL and PEL sites. PC1 and PC2 stand for Principal Components 1 and 2, respectively.

Descriptive statistics and population structure

There was no evidence of null alleles or stuttering errors in the microsatellites studied. Deviations from the HWE or linkage disequilibrium were not consistent for any particular locus and, therefore, all loci were included in the analyses. Relatedness values

also showed no evidence for a high aggregation of related individuals in the collected samples (Annex 2).

The assignment analysis performed with the FLOCK software did not determine the exact k value, but it indicated that $k \geq 3$ in the Maipo River basin. The GENELAND software found evidence of five populations in 60 % of the runs performed. The run showing the highest average log posterior probability confirmed that $k = 5$. An individual assignment to the detected populations was performed for several runs, but no coherence among these was found. The global F_{ST} in the Maipo River basin was 0.089 ($P < 0.0001$), indicating that at least one group of sampled silversides was different from the total of sites analyzed. Pairwise analysis suggested the presence of five different populations and showed F_{ST} values ranging from 0.009 to 0.143 (Table 1). Considering the results of these three methods, the genetic structure retained for subsequent analyses considered the presence of five isolated groups within the Maipo River basin. Taking into account the results of both the BCWQI and the PCA, through which it was possible to define non-polluted and polluted sites, two populations of *B. microlepidotus* were found to inhabit polluted sites (MEL and PEL) and three populations inhabited non-polluted sites (MA-SFM, IM-PN, PU) (Fig. 1). Associated with these results is that the Mantel test showed no evidence of a relationship between F_{ST} and geographical distance among the sampling sites ($r = -0.0862$, $P = 0.61$).

Table 1 Pairwise F_{ST} and associated P values obtained after 10000 permutations.

	SFM[1]	MA[1]	PU[1]	PN[1]	PEL[2]	MEL[2]	IM[1]
SFM[1]		0.009	0.063***	0.143***	0.081***	0.075***	0.1***
MA[1]			0.052***	0.143***	0.038**	0.094***	0.137***
PU[1]				0.103***	0.052***	0.096***	0.121***
PN[1]					0.087***	0.084***	0.024
PEL[2]						0.086***	0.136***
MEL[2]							0.057**

**: $P < 0.01$; ***: $P < 0.001$

[1]Non-polluted sites according to BCWQI.

[2]Polluted sites according to BCWQI.

Using the five populations classified to polluted and non-polluted environments, the comparison of N_A, AR, and GD did not show significant differences ($P = 0.21$, $P = 0.35$, and $P = 0.47$, respectively). Furthermore, no significant evidence of a bottleneck was detected for any of the populations (PU $P = 0.742$, MA-SFM $P = 0.461$, MEL $P = 0.054$, PEL $P = 0.469$, and IM-PN $P = 0.844$).

Historical and contemporary gene flow and effective population size estimates

Current migration (Table 2) showed high self-recruitment at IM-PN and PEL (> 90 %), with lower values found at the other sites. An interesting result was that the PEL site has the highest migration rate as compared to all other sites, and this pattern was not observed in the context of historical migration (Table 3). Moreover, the values of historical migration were lower than those observed for current migration. Historical and current N_em were between 0.309 - 4.135 and 0.224 - 7.526, respectively (Annex 3 and

4). This same pattern was observed in the current N_em, where the PEL site had the highest N_em exported to all other sites.

Table 2 Current migration rate estimated for populations of *B. microlepidotus* in the Maipo River basin. Each value represents a mean of ten different runs. Standard deviations are shown in parentheses.

From/To	MEL	IM-PN	MA-SFM	PU	PEL
MEL	0.688 (4×10^{-4})	0.014 (3×10^{-4})	0.007 (6×10^{-5})	0.012 (6×10^{-5})	0.012 (1×10^{-4})
IM-PN	0.026 (2×10^{-4})	0.931 (2×10^{-4})	0.008 (5×10^{-5})	0.031 (2×10^{-4})	0.015 (3×10^{-5})
MA-SFM	0.033 (4×10^{-4})	0.018 (1×10^{-4})	0.856 (1×10^{-3})	0.089 (1×10^{-3})	0.018 (9×10^{-5})
PU	0.013 (1×10^{-4})	0.012 (9×10^{-4})	0.035 (2×10^{-3})	0.746 (2×10^{-3})	0.027 (5×10^{-4})
PEL	0.239 (7×10^{-4})	0.024 (1×10^{-4})	0.094 (1×10^{-3})	0.122 (2×10^{-3})	0.928 (5×10^{-4})

Table 3 Historical migration rate estimated for populations of *B. microlepidotus* in the Maipo River basin. Each value represents a mean of three different runs. Standard deviations are shown in parentheses.

From/To	MEL	IM-PN	MA-SFM	PU	PEL
MEL		0.008 (2×10^{-3})	0.017 (2×10^{-3})	0.007 (4×10^{-3})	0.024 (3×10^{-3})
IM-PN	0.006 (1×10^{-3})		0.002 (1×10^{-3})	0.01 (6×10^{-3})	0.018 (7×10^{-3})
MA-SFM	0.011 (6×10^{-3})	0.007 (3×10^{-3})		0.006 (3×10^{-3})	0.009 (7×10^{-3})
PU	0.006 (6×10^{-3})	0.009 (5×10^{-3})	0.005 (2×10^{-3})		0.002 (2×10^{-3})
PEL	0.013 (8×10^{-3})	0.011 (6×10^{-3})	0.014 (8×10^{-3})	0.008 (3×10^{-3})	

An increase in the current migration rate for all populations of the Maipo River basin was detected with respects to the historical migration rate (historical m = 0.0098 ± 0.0053, current m = 0.0425 ± 0.0547, Fig. 3a), but this difference was not observed for N_em (historical N_em = 1.545 ± 1.123, current N_em = 1.529 ± 1.649, Table 4, Fig. 3b). This could be explained by the current low N_e when compared to the historical value

(historical N_e = 178.51 ± 103.72, current N_e = 48.171 ± 33.48, V = 55, P < 0.01, Fig. 4). On the other hand, no statistical differences were detected between current I/E ratios (for both m and N_em) from the polluted and non-polluted populations (Table 4). Additionally, no changes were detected between the historical and current I/E ratios (for both m and N_em) of the non-polluted populations. The same applies to the polluted populations (Table 4). Thus, no effects of pollution on migration rates or effective migrants per generation of *B. microlepidotus* were detected.

Table 4 Results of the permutation ANOVA test. Top: results for migration rate (m) and effective migrants per generation (N_em) comparing current and historical estimations. Middle: comparisons for m and N_em of the current immigration/emigration ratio (I/E ratio) between polluted and non-polluted populations. Bottom: historical vs current I/E ratio for m and N_em for non-polluted and polluted populations, separately. Bold values indicate significant differences.

	F value	P value
Migration: historical vs current		
All populations (m)	6.767	**0.001**
All populations (N_em)	0.001	0.984
I/E ratio: non-polluted vs polluted populations		
Current (m)	0.778	0.533
Current (N_em)	0.145	0.772
I/E ratio: historical vs current		
Non-polluted populations (m)	0.203	0.648
Polluted (m)	1.240	0.352
Non-polluted populations (N_em)	2.250	0.125
Polluted (N_em)	1.843	0.147

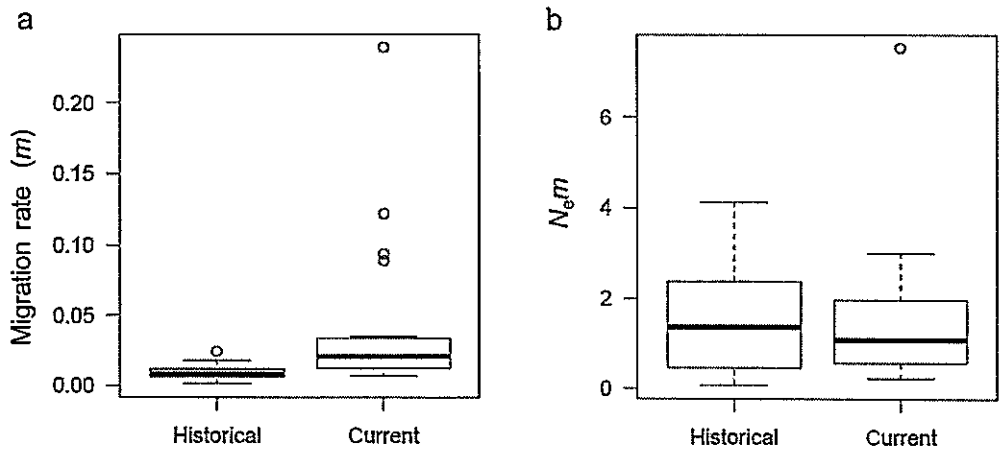


Figure 3. Boxplot of the historical and current a) migration rates (m) and b) effective migrants per generation (N_em). The box represents the first and third quartiles, and the whiskers are the minimum and maximum values. The band inside the box is the median (second quartile), and the points represent the outliers.

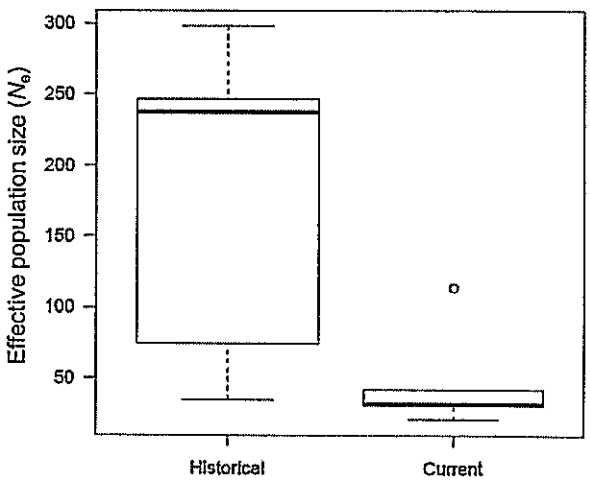


Figure 4 Boxplot of the historical and current effective population size (N_e) estimated for populations of *B. microlepidotus* in the Maipo River basin. The box represents the first and third quartiles, and the whiskers are the minimum and maximum values. The band inside the box is the median (second quartile), and the points represent the outliers.

DISCUSSION

Pollution is currently a main factor affecting natural populations in freshwater systems, as has been documented for different taxa (e.g., Dudgeon et al. 2006; Strayer and Dudgeon 2010), and the effects caused by pollution are very significant in rivers surrounded by large cities around the world. Freshwater quality is mainly determined by natural processes, such as precipitation, weathering processes, and soil erosion, and by anthropogenic influences, such as the use of water resources as well as industrial and agricultural activities (Singh et al. 2005), both affecting the habitat and, finally, the biota.

In this study, the water quality of the Maipo River basin was determined with the BCWQI and showed sites ranging from the fair to excellent categories. In addition to this index, the PCA indicated that TDS, EC, Cl^-, and SO^{2-}_4 explained 51 % of the variance and showed a segregation of the sites that were ranked within the fair category of water quality. From these two results, the sites with an excellent or good ranking (PU, IM, PN, MA, SFM) were defined as non-polluted, and the two sites with a fair ranking (PEL, MEL) were defined as polluted. The two sites classified as polluted showed a TDS over 620 mg/L, which is almost three times greater than the 261.96 mg/L average measured at seven sites of the Gomti River, which is considered one of the most polluted rivers in India (Singh et al. 2005). Furthermore, the two polluted sites of the Maipo River showed values of EC > 1200 µS/cm. These values were greater than the 1110 µS/cm detected for the influent of the Keiskammahoek Sewage Treatment Plant of the South African Keiskamma River (Morrison et al. 2001) or the 597.5 µs/cm detected in

the Brazilian Piracicaba River basin, which is located in a highly developed area of Brazil that produces about 10 % of the internal product and 85 % of the urban sewage is dumped into rivers and streams without any treatment (Daniel et al. 2001). It is important to note that high values of EC could be related to wastewater effluents, storm water drainage, or industrial effluent discharges.

Based on our sampling at a total of seven sites in the Maipo River Basin, five populations of the silverside *B. microlepidotus* were detected. According to the Mantel test this differentiation was not explained by geographical distance, and taking into account that there are no apparent barriers among the sampling sites (e.g., dams), this differentiation could be associated with the high anthropic perturbation present in this basin. Despite the high levels of intervention in this basin, relatively clean sites were found. This is probably related to the presence of groundwater that joins perturbed water flowing to the sea, as was the case at the MA, PN, and IM sites. It is important to note that the affluent at the IM site is located within an agricultural area, whose owners are known to take care of the water quality. In the case of the PU and SFM sites, both were located away from the large city, and geographical isolation tends to produce a low human impact, hence resulting in areas of low pollution.

This study did not find evidence of pollution affecting silverside populations, and two of the detected populations inhabited polluted sites. In contrast to the results drawn by Theodorakis et al. (2006), the present study did not find evidence of pollution influencing levels of gene flow for *B. microlepidotus* in the Maipo River basin.

However, a global increase in the current migration rate as compared to the historical rate was observed, but this was not found to be related to pollution and was not enough to homogenize all populations within the Maipo River basin. Moreover, no differences in N_A, AR, and GD between the non-polluted and polluted areas were detected. Klerks et al. (1997) found a similar pattern, where no differences in the frequencies of allozyme genotypes and heterozygosity levels among *Gobionellus boleosoma* (Jordan and Gilbert) were detected between a polluted area and a control site. For the non-migratory fish *Fundulus heteroclitus* (Linnaeus), AFLP markers distinguished populations along a gradient of polychlorinated biphenyl contamination. For this fish, genome-wide diversity was also preserved across study sites, and this was due to the large effective population sizes and/or because the mechanism for genetic adaptation to these contaminants affected only a small number of loci (McMillan et al. 2006). Regarding *B. microlepidotus*, however, the explanation could not be a large effective population size at the polluted populations because they showed a low current N_e (MEL = 31.5, PEL = 42.3), which was, in general, lower than historical values. Thus, it is probable that pollution does not affect N_A, AR, and GD in *B. microlepidotus*, as confirmed by the absence of bottlenecks in the populations. However, the *B. microlepidotus* population at the MEL site showed a marginal P value ($P = 0.054$), which was probably due to small sample size (see Johnson 1999). Thus, further studies in this population could be conducted in order to clarify the effects of pollution and the changes in population size over time. Furthermore, *B. microlepidotus* inhabits lakes and rivers in Chile from 28° up to 39°S, and this gradient includes various environments, which also indicates that this species is able to tolerate changes and different freshwater environments. Taken

together, all of the obtained results indicate that this species is tolerant to the types of pollution found in the Maipo River basin.

Despite this, the significant reduction of the current N_e and the endangered status of this species make it urgent to develop a management plan for its conservation in disturbed areas. Based on our analyses, we consider the conservation of some sites away from the large city (e.g., PU and SFM) is a valuable strategy towards creating a network of populations that could interact with migrants and conserve the gene diversity of this species in the basin. Additionally, the polluted site PEL could be an important area to conserve due to the number of progenies exported from here to the different populations within in the Maipo River basin. Furthermore, we observed a greater number of young individuals in this area as compared to the other sites visited in the Maipo River basin, suggesting that it is an important site for this species in terms of reproduction, which also makes it an essential site for future studies.

The present report did not find evidence to support the claim that pollution affects gene diversity. Additionally, sites with and without pollution did not show differences in current migration rates when compared to historical migration rates. However, the present findings contributes to the understanding that not all native species respond in the same way to pollution and that case by case studies are necessary to create an adequate management plan for specific species.

Finally, the observed tolerance to polluted environments of *B. microlepidotus* could be explained by the use of non-neutral markers. Williams and Oleksiak (2008) determined that 1 to 6 % of the AFLP loci used in a study of *Fundulus heteroclitus* populations exposed to chronic contamination could be under selection or linked to areas of the genome that are under selection due to pollution. The expression of functional genes plays an important role in physiological acclimation and evolutionary adaptation to diverse environments, which would be the next step in studying how *B. microlepidotus* can adapt to polluted areas.

Acknowledgements

We thank L. Eaton for reviewing the English version of the manuscript, JP Oyanedel for comments on an earlier version and C. Quezada-Romegialli for helping us in the field work. Thanks to C. Poveda, V. Nuñez, G. Laree and L. Mauro for chemical analyses. DV thanks the financial support of Fondecyt 1100341, Proyecto Enlace Universidad de Chile 2013, Basal Grant PFB 023 and ICM P05-002; CVR thanks Conicyt Doctoral Fellowship 21090188 and doctoral thesis fellowship 24121005.

REFERENCES

Abell R, Allan JD, Lehner B (2007) Unlocking the potential of protected areas for freshwaters. Biol Conserv 134:48–63.

Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Res 25:4692–4693.

Bahamondes I, Soto D, Vila I (1979) Hábitos alimentarios de las especies de Atherinidae del Embalse Rapel. Medio Ambiente 4:3–18 (in Spanish).

Baker RJ, Bickham AM, Bondarkoy M, Gascjak SP, Matson CW, Rodgers BE, Wickcliffe JK, Chesser RK (2001) Consequences of polluted environments on population structure: the bank vole (*Clethrionomys glareolus*) at Chernobyl. Ecotoxicology 10:211–216.

BCWQI (1996) The British Columbia Water Quality Index, Environmental Protection Division. Ministry of Environment Government of British Columbia. http://www.env.gov.bc.ca/wat/wq/BCguidelines/indexreport.html. Accessed 19 February 2014.

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics 152:763–773.

Beheregaray LB, Sunnucks P (2000) Microsatellite loci isolated from *Odontesthes argentinensis* and the *O. perugiae* species group and their use in the other South American silverside fish. Mol Ecol 9:629–644.

Belfiore NM, Anderson SL (2001) Effects of contaminants on genetic patterns in aquatic organism: a review. Mutat Res 489:97–122.

Belkhir K, Borsa P, Chikki L, Goudet J, Bonhomme F (2000) GENETIX version 4.02. Logiciel sous WindowsMT pour la Genétique des Populations. Laboratoire Génome et Populations, CNRS UPR 9060, Université de Montpellier II, Montpellier, France (in French).

Belkhir K, Castric V, Bonhomme F (2002) Identix, a software to test for relatedness in a population using permutation methods. Mol Ecol Notes 2:611–614.

Bickham JW, Sandhu S, Hebert PDN, Chikhi L, Athwal R (2000) Effects of chemical contaminants on genetic diversity in natural populations: implications for biomonitoring and ecotoxicology. Mutat Res 463:33–51.

Campos H, Dazarola G, Dyer B, Fuentes L, Gavilan J, Huaquin L, Martinez G, Melendez R, Pequeño G, Ponce F, Ruiz V, Siefeld W, Soto D, Vega R, Vila I (1998) Categorías de conservación de peces nativos de aguas continentales de Chile. Boletin del Museo Nacional de Historia Natural 47:101–122 (in Spanish).

Clesceri LS, Greenberg AE, Eaton AD (2005) Standard Methods for the examination of water and wastewater, 20th edn. American Public Health Association, American Water Works Association, Water Environment Federation, Washington, DC.

Comte S, Vila I (1992) Spawning of *Basilichthys microlepidotus* (Jenyns). J Fish Biol 41:971–981.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014.

Daniel MHB, Montebelo AA, Bernardes MC, Ometto JPHB, Camargo PB, Krusche AV, Ballester MV, Victoria RL, Martinelli LA (2001) Effects of urban sewage on dissolved oxygen, dissolved inorganic and organic carbon, and electrical conductivity of small streams along a gradient of urbanization in the Piracicaba river basin. Water Air Soil Poll 136:189–206.

DiBattista JD (2008) Patterns of genetic variation in anthropogenically impacted populations. Conserv Genet 9:141–156.

Dirección general de aguas (2004) Diagnostico y clasificación de los cursos y cuerpos de agua según objetivos de calidad, cuenca del Rio Maipo. Ministerio de Obras Públicas, Gobierno de Chile (in Spanish).

Domínguez-Domínguez O, Boto L, Alda F, Pérez-Ponce de León G, Doadrio I (2007) Humans impacts on drainages of the mesa central Mexico, and its genetics effects on an endangered fish, *Zoogoneticus quitzeoensis*. Conserv Biol 21:168–180.

Duarte W, Feito R, Jara C, Moreno C, Orellana AE (1971) Ictiofauna del sistema hidrográfico del río Maipo. Boletin del Museo Nacional de Historia Natural 32:227–268 (in Spanish).

Duchesne P, Turgeon J (2012) FLOCK provides reliable solutions to the "Number of populations" problem. J Hered 103:734–743.

Dudgeon D, Arthington AH, Gessner MO, Kawabata Z-I, Knowler DJ, Lequeve C, Naiman RJ, Prieur-Richard A-H, Soto D, Stiassny MLJ, Sullivan CA (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. Biol Rev 81:163–182.

Dukes JS, Mooney HA (1999) Does global change increase the success of biological invaders? Trends Ecol Evol 14:135–139.

Durrant CJ, Stevens JR, Hogstrand C, Bury NR (2011) The effect of metal pollution on the population genetic structure of brown trout (*Salmo trutta* L.) residing in the River Hayle, Cornwall, UK. Environ Pollut 159: 3595–3603.

Fagúndez J (2013) Heathlands confronting global change: drivers of biodiversity loss from past to future scenarios. Ann Bot-London 111:151–172.

Fahrig L (1997) Relative effects of habitat loss and fragmentation on population extinction. J Wildl Manage 61:603–610.

Fracácio R, Verani NF, Gaeta Espíndola EL, Rocha O, Rigolin-Sá O, Andrade CA (2003) Alterations on growth and gill Morphology of *Danio rerio* (Pisces, Ciprinidae) exposed to the toxic sediments. Braz Arch Biol Techn 46:685–695.

Goudet J (1995) FSTAT: a computer program to calculate F-statistics. J Hered 86:485–486.

Guillot G, Mortimer F, Estoup A (2005) GENELAND: a computer package for landscape genetics. Mol Ecol Notes 5:712–715.

Jackson RB, Carpenter SR, Dahm CN, McKnight DM, Naiman RJ, Postel SL, Running SW (2001) Water in changing world. Ecol Appl 11:1027-1045.

Johnson DH (1999) The insignificance of statistical significance testing. J Wildl Manage 63:763–772.

Jones A, Rosenqvist G, Berglund A, Avise JC (1999) Clustered microsatellite mutations in the pipefish *Syngnathus typhle*. Genetics 152:1057–1063.

Klerks PL, Leberg PL, Lance RF, McMillin DJ, Means JC (1997) Lack of development of pollutant-resistance or genetic differentiation in darter gobies *(Gobionellus boleosoma*) inhabiting a produced-water discharge site. Mar Environ Res 44:317–395.

Koshimizu E, Strüssmann CA, Tejedor ED, Okamoto N, Fukuda H, Sakamoto T (2009) Development of polymorphic microsatellite *loci* for two Atherinopsid fishes, pejerrey (*Odontesthes bonariensis*) and Patagonian pejerrey (*O. hatcheri*). Mol Ecol Resour 9:1460–1466.

Mackiewicz M, Fletcher DE, Wilkins SD, DeWoody J, Avise J (2002) A genetic assessment of parentage in a natural population of dollar sunfish (*Lepomis marginatus*) based on microsatellite markers. Mol Ecol 11:1877–1883.

McMillan AM, Bagley MJ, Jackson SA, Nacci DE (2006) Genetic diversity and structure of an estuarine fish (*Fundulus heteroclitus*) indigenous to sites associated with a highly contaminated urban harbor. Ecotoxicology 15:539–548.

Mora C, Metzger R, Rollo A, Myers RA (2007) Experimental simulations about the effects of overexploitation and habitat fragmentation on populations facing environmental warming. Proc R Soc B-Biol Sci 274:1023–1028.

Morrison G, Fatoki OS, Persson L, Ekberg A (2001) Assessment of the impact of point source pollution from the Keiskammahoek Sewage Treatment Plant on the Keiskamma River - pH, electrical conductivity, oxygen- demanding substance (COD) and nutrients. Water SA 27:475–480.

Muñoz BM (2007) Cambios en la riqueza íctica del rio Maipo. Tesis Biología con Mención en Medio Ambiente, Facultad de Ciencias, Universidad de Chile (in Spanish).

Muñoz P, Quezada-Romegialli C, Vila I, Véliz D (2011) Cross-amplification of microsatellites from the Atherinopsidae *Odontesthes perugiae* and *Odontesthes argentinensis* to Chilean silversides of the genus *Odontesthes* and *Basilichthys*. Gayana 75:182–186.

Murdoch MH, Hebert PDN (1994) Mitochondrial DNA diversity of brown bullhead from contaminated and relatively pristine sites in the Great Lakes. Environ Toxicol Chem 13:1281–1289.

Pandey S, Parvez S, Ansari RA, Ali M, Kaur M, Hayat F, Ahmad F, Raisuddin S (2008) Effects of exposure to multiple trace metals on biochemical, histological and ultrastructural features of gills of a freshwater fish, *Channa punctata* Bloch. Chem-Biol Interact 174:183–192.

Pardo R, Vila I, Bustamante R (2008) Demographic response of *Stratiodrilus aeglaphilus* (Anelida, Histriobdellidae) to organic enrichment: experimental assessment. Hydrobiologia 598:11–16.

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. Evolution 43:258–275.

Quezada-Romegialli C, Fuentes M, Veliz D (2010) Comparative population genetics of *Basilichthys microlepidotus* (Atheriniformes: Atherinopsidae) and *Trichomycterus areolatus* (Siluriformes: Trichomycteridae) in north central Chile. Environ Biol Fish 89:173–186.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Sala OE, Chapin III FS, Armesto JJ, Berlw E, Bloomfield J, Dirzo R, Huber-Sanwald E, Huenneke LF, Jackson RB, Kinzing A, Leemans R, Lodge DM, Mooney HA, Oesterheld M, Poff NL, Sykes MT, Walker BH, Walker M, Wall DH (2000) Global biodiversity scenarios for the year 2100. Science 287:1770–1774.

Singh KP, Malik A, Sinha S (2005) Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—a case study. Analytica Chimica Acta 538:355–374.

Strayer DL, Dudgeon D (2010) Freshwater biodiversity conservation: recent progress and future challenges. J N Am Benthol Soc 29:344–358.

Strickland JDH, Parsons TR (1968) Determination of dissolved oxygen in a practical handbook of seawater analysis. J Fish Res Board Can 167:71–75.

Tallmon DA, Koyuk A, Luikart GH, Beaumont MA (2008) ONeSAMP: a program to estimate effective population size using approximate Bayesian computation. Mol Ecol Resour 8:299–301.

Theodorakis CW (2003) Establishing causality between population genetic alterations and environmental contamination in aquatic organisms. Hum Ecol Risk Assess 9:37–58.

Theodorankis CW, Lee KL, Adams M, Law B (2006) Evidence of altered gene flow, mutation rate, and genetic diversity in redbreast sunfish from a pulp-mill-contaminated river. Environ Sci Technol 40:377–386.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Microchecker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538.

van Straalen NM, Timmermans MJTN (2002) Genetic variation in toxicant stressed populations: an evaluation of the "genetic erosion" hypothesis. Hum Ecol Risk Assess 8:983–1002.

Véliz D, Catalán L, Pardo R, Acuña P, Diaz A, Poulin E, Vila I (2012) The genus *Basilichthys* (Teleostei : Atherinopsidae) revisited along its Chilean distribution range (21° to 40°S) using variation in morphology and mtDNA. Rev Chil His Nat 85:49-59.

Vila I, Pardo R, Dyer B, Habit E (2006) Peces límnicos: diversidad, origen y estado de conservación. In: Vila I, Veloso A, Schlatter R, Ramírez C (eds) Macrófitas y

vertebrados de los sistemas límnicos de chile, 1st edn. Editorial Universitaria, Santiago, pp 73–102 (in Spanish).

Wasko AP, C Martins, Oliveira C, Foresti F (2003) Non-destructive genetic sampling in fish: An improved method for DNA extraction from fish fins and scales. Hereditas 138:161–165.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370.

Weis JS, Smith G, Zhou T, Santiago-Bass C, Weis P (2001) Effects of contaminants on behavior: Biochemical mechanisms and ecological consequences. Bioscience 51:209–217.

Williams LM, Oleksiak MF (2008) Signatures of selection in natural populations adapted to chronic pollution. BMC Evol Biol 8:282–293.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163:1177–119.

World Water Assessment Programme (2009) The United Nations World Water Development Report 3: Water in a Changing World. UNESCO, Paris and Earthscan, London.

CHAPTER 2

SIGNATURES OF DIRECTIONAL AND BALANCING SELECTION IN THE SILVERSIDE *BASILICHTHYS MICROLEPIDOTUS* (TELEOSTEI: ATHERINOPSIDAE) INHABITING A POLLUTED RIVER

Caren Vega-Retter[1,2], Irma Vila[1], David Véliz[1,2]

[1]Departamento de Ciencias Ecológicas, [2]Instituto de Ecología y Biodiversidad, Universidad de Chile, Las Palmeras 3425, Ñuñoa, Santiago, Chile. e-mail: dveliz@uchile.cl, Telephone: 56229787314, Fax: 56222727363

Running title: Selection in a fish inhabiting a polluted river

Submitted to Journal of Evolutionary Biology

ABSTRACT

Currently environmental pollution is one of the most important factors affecting natural populations and acting as a strong selective pressure. Therefore, identifying genes and their variants (alleles) implied in population survival within contaminated areas is a relevant issue. Freshwater systems are among those that have been most impacted by pollution. The Maipo River is one of the most polluted basins in Chile, surrounded by 40% of the human population of the country. There are five populations of the endemic silverside *Basilichthys microlepidotus* inhabiting this river, two in polluted areas and three in non-polluted areas. The goal of this study was to determine the proportion of loci under directional and balancing selection related to pollution in *B. microlepidotus*. To this end, a genome scan (AFLP markers) was performed comparing between fish located in polluted and non-polluted areas and between fish inhabiting polluted sites. Ten loci (6.7% of the total loci) were identified as candidate loci under selection; six (4.0%) under directional selection and four (2.7%) under balancing selection. This study contributes to demonstrate that pollution could be implicated in selection even within a basin. As far as we know, this is the first study to date that has detected loci under balancing selection related to pollution, indicating that pollution could be responsible for maintaining polymorphism in these loci.

Key words: balancing selection, directional selection, pollution, AFLP.

INTRODUCTION

Understanding the genetic basis of adaptation is currently a fundamental issue in evolutionary biology. Alterations in both abiotic and biotic environmental conditions may change the direction, strength and form of selection, thus becoming an issue of importance in the context of global change (Reusch and Wood 2007). Diverse approaches have been developed in the area of population genomics in order to determine the distribution of genomic variation within and between populations, namely candidate genes, population transcriptomics, quantitative trait loci and genome scans (Reusch and Wood 2007; Nielsen et al. 2009). A genome scan allows the identification of loci that show abnormal levels of structuring and/or polymorphism, potentially as a result of selection (Luikart et al. 2003; Beaumont 2005; Storz 2005).

The loci recognized in a genome scan are, therefore, candidates for genes or genomic regions being under either balancing or directional selection. Loci under directional selection are expected to decrease allele diversity within populations and to increase differentiation among populations compared to neutral loci. On the other hand, balancing selection homogenizes allele frequencies among populations, suggesting that balancing selection would be responsible for maintaining stable polymorphisms in time (Nielsen 2005; Charlesworth 2006; Veliz et al. 2006). Thus genomic regions showing such patterns of genetic diversity could be considered as candidates for containing loci involved in evolutionary processes (Schlötterer 2003).

A useful marker in genome scans has been the Amplified Fragment Length Polymorphism (AFLP); this technique can easily be applied to non-model organisms, producing hundreds of loci widely distributed across the genome (Meudt and Clarke 2007). This approach has led to the identification of adaptive divergence in different taxa. Campbell and Bernatchez (2004), for example, showed that between 1.4% and 3.2% of the scored AFLP loci were linked to genes implicated in the adaptive radiation of the whitefish, *Coregonus clupeaformis* (Mitchill), while Bonin *et al.* (2006) detected eight candidate loci potentially involved in adaptation to altitude in *Rana temporaria* (Linnaeus). In addition, Henry and Russello (2013) reported that selection drives divergence in a climate-change-sensitive mammal (*Ochotona princeps* Richardson) using these markers.

Currently, environmental pollution is one of the most important factors affecting natural populations and acting as a strong selective pressure. Thus identifying the proportion of genes and their variants (alleles) implied in organism survival within contaminated areas is a relevant issue and constitutes the first step towards a global understanding of the evolutionary responses to environmental pressures. Paris et al. (2010) found signatures of positive selection in 5 loci (3.2% of loci studied) of the mosquito *Aedes rusticus* (Rossi) after treatment with insecticide. In the case of plants, four loci were found under divergent selection in *Arabidopsis halleri* (Linnaeus), hence they could be considered as loci for general pollution adaptation (Meyer et al. 2009). In the fish *Fundulus heteroclitus* (Linnaeus), Williams and Oleksiak (2008) found evidence that three populations inhabiting contaminated sites showed from 1% to 6% of the loci under

selection or loci related to the genome area under selection. The results in all these studies suggest rapid adaptive evolution in response to environmental pollution.

Freshwater systems are among the most affected by pollution, which is the principal source of water quality degradation in the world (World Water Assessment Programme 2009). One of the most polluted basins in Chile is the Maipo River basin, where approximately 40% of the Chilean population lives, that is, almost 6.7 million inhabitants according to the most recent census in 2012. This basin has experienced water quality deterioration and eutrophication (Pardo et al. 2008), mostly as a result of organic matter from untreated sewage. Additionally, this basin contains the largest number of factories in the country as well as mining facilities in the Andes Range, which add heavy metals into this river (Dirección General de Aguas 2004d).

Studies on fish diversity in the Maipo River basin have shown a significant reduction in richness and species abundance during the last 30 years (Muñoz 2007), which may be due to habitat degradation and pollution. One species that has shown an important decrease in its abundance is the silverside *Basilichthys microlepidotus* (Jenyns), an atherinopsid endemic to Chile that is considered to be endangered (Campos *et al.*, 1998). This species inhabits lakes and rivers from 28° to 39°S (Véliz et al. 2012), feeds on insect larvae, small invertebrates, filamentous algae and detritus (Duarte et al. 1971; Bahamondes et al. 1987) and reproduces from August to January (Comte and Vila 1992). A previous study performed in the Maipo River basin using microsatellite loci showed the presence of populations of *B. microlepidotus* inhabiting polluted areas of the river and of others inhabiting non-polluted sites (Vega-Retter et al. 2014). Thus this

species and the basin that it inhabits arises as an interesting model system to study the selective effect of the pollution in the genome.

For this analysis we used an AFLP-based genome scan to determine the proportions of loci under directional or balancing selection between *B. microlepidotus* located in polluted and non-polluted areas and between individuals located in polluted sites, respectively. This comparison is the first step in our knowledge of selection related to pollution in this fish.

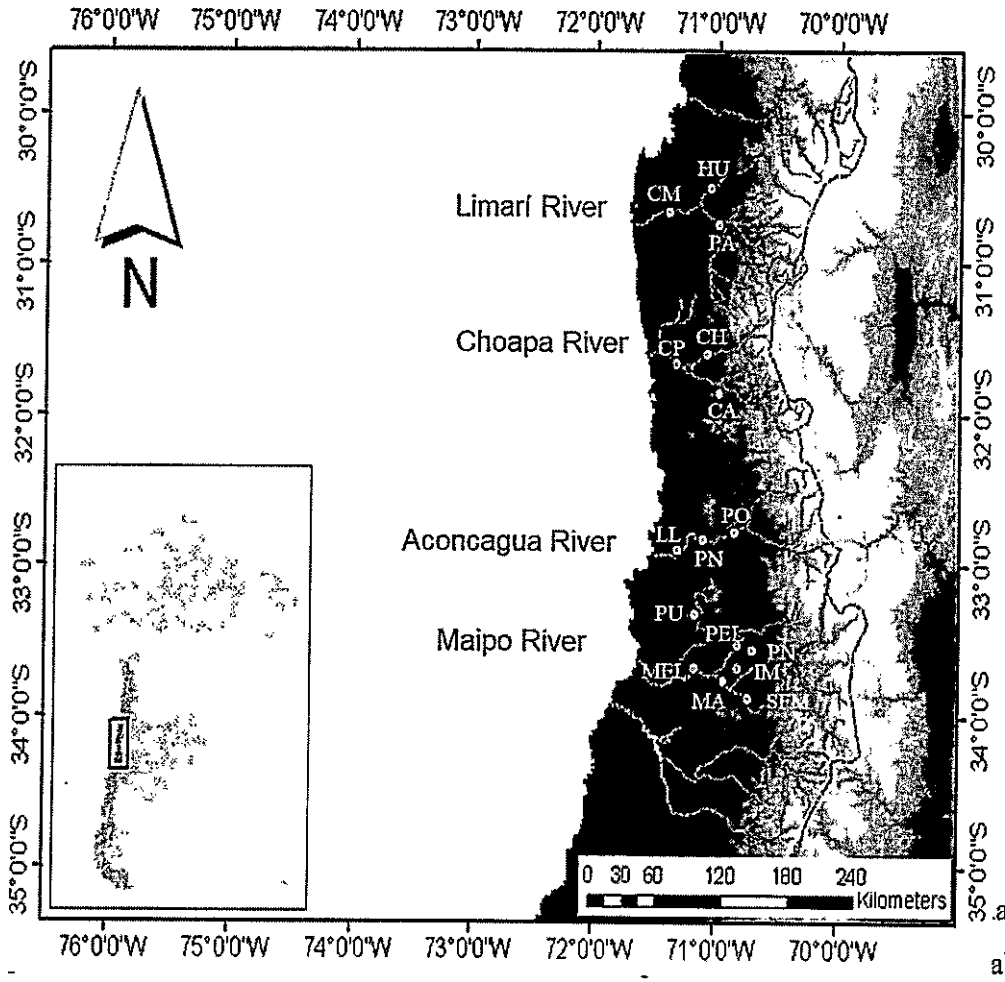
MATERIAL AND METHODS

Sampling sites and collection

In order to detect loci under both directional and balancing selection related to pollution, seven sites were sampled in the Maipo River. A total of 166 specimens were collected by using an electrofishing device; a mean of 23 individuals per site (Fig. 1). Vega-Retter et al. (2014) found five different populations of silversides in this basin, two located in sites classified as highly polluted (PEL and MEL) and three populations inhabiting five sites classified as non-polluted (PU, IMPN and MASFM), according to the British Columbia Water Quality Index (BCWQI 1996). This population structure was considered for the present study.

Three other basins were sampled as control sites (i.e., non-polluted sites), namely the Limarí, Choapa and Aconcagua rivers located in north-central Chile (Fig. 1). These basins have been impacted by lower degrees of industrial development compared to the Maipo River Basin. Moreover, the human population surrounding these rivers is also considerably less, with 66,000, 152,000 and 430,000 people living in the Limarí, Choapa and Aconcagua River basins, respectively (Dirección General de Aguas 2004a, b, c), compared to approximately 6.7 million habitants living around the Maipo River basin according to the 2012 census. Thirty *B. microlepidotus* individuals were sampled by means of electrofishing from three different sites within each basin, thus totaling 270

specimens. The tissue samples were obtained by dissecting a small piece of caudal fin stored in 99% ethanol.



.a and aballo Muerto; within Choapa basin: CH: Chillepin; CA: Camisas; CP: Choapa Pueblo; in the Aconcagua basin: PO: Pocuro; PN: Panuehue; LL: LlaLlay; in the Maipo basin: PU: Puangue; PEL: Pelvin; PN: Peñaflor; IM: Isla de Maipo; MEL: Melipilla; MA: Maipo; SFM: San Francisco de Mostazal.

DNA extraction and AFLP amplification

DNA extraction was performed using the salt extraction method (Aljanabi & Martinez, 1997) and a nanospectrophotometer was used to quantify DNA concentration. The AFLP procedure was performed following Vos et al. (1995) and the restriction enzymes EcoRI and MseI were used for genomic DNA digestion. The following primers were used for preselective PCR: EcoRI (GACTGCGTACCAATTCA) and MseI (GATGAGTCCTGAGTAAC), followed by a PCR of four combinations of selective primers: *Eco*RI–ACT/ *Mse*I–CTT, *Eco*RI–ACC/*Mse*I–CTC, EcoRI–ACA/MseI–CAC and MseI –CTA/ EcoRI–AAC, with the *Mse*I primer containing the fluorescent dye. PCR products were loaded in ABI capillary sequencers (Applied Biosystems, Foster City, CA, USA) in Macrogen Inc. (South Korea) to visualize the fragments. The LIZ500 was used as internal size standard. GENEMARKER software was used to analyze the fragments; only clean peaks of between 75 and 450 bp were used in the analysis, coded as 1/0 to indicate the presence/absence of each fragment in each individual. To avoid possible scoring errors, only loci with more than 1% polymorphism were used.

Detection of AFLP loci under selection

In order to detect loci under directional and balancing selection, the F_{ST} outlier method was performed using the following two software programs: DFDIST (Beaumont and Nichols 1996) and BAYESCAN (Foll and Gaggiotti 2008). DFDIST, a variant of the FDIST program, compares the empirical F_{ST} values with a null distribution derived from a coalescence simulation, determining the likelihood of the F_{ST} values to be greater or lower than what is observed under neutrality. DFDIST was implemented in the

MCHEZA program (Antao and Beaumont 2011), which applies a multi-test correction based on false discovery rate (FDR) to avoid overestimation of the outlier loci (Caballero et al. 2008). The runs were conducted with the following settings: 80,000 simulations; theta, beta-a, beta-b at the default values of 0.1, 0.25 and 0.25, respectively; the significant effects were considered with 95% confidence level. The initial mean dataset F_{ST} is often not neutral, since the selected loci are included in the computation. Then the function "Neutral mean F_{ST}" was used to determine a first candidate subset of selected loci in order to remove them from the computation of the neutral F_{ST}. The function "Force mean F_{ST}" was chosen to simulate a precise mean F_{ST} by running a bisection algorithm over repeated simulations. Loci with a significant *P*-value after using a FDR threshold of 0.1 (10%) were considered as candidate loci; F_{ST} values above or below the expected values were considered to be under directional or balancing selection, respectively.

The second analysis was performed with the BAYESCAN software, which directly estimates the posterior probability of a locus being under selection. For this, BAYESCAN uses the differences in allele frequencies between populations. In brief, this method is an extension of that proposed by Beaumont and Balding (2004), based on a logistic regression model in which each logit value of genetic differentiation F_{ST} (i, j) for locus i in population j is decomposed as a linear combination of the coefficients of the logistic regression, α_i and β_j, corresponding to a locus effect and to a population effect, respectively. Departure from neutrality at a given locus is assumed when the locus-specific component is necessary to explain the observed pattern of diversity ($\alpha \neq$

0); with $\alpha > 0$ suggesting directional selection, and $\alpha < 0$ the presence of balancing selection. BAYESCAN was run with 20 pilot runs, a burn-in and sample size of 10,000 and thinning interval of 50. Loci were considered as candidates for being under selection if their $\log_{10}$ Bayes Factor scores were ≥ 0.5 ("Substantial" on Jeffrey's scale of evidence) (Foll and Gaggiotti 2008). In addition, FDR was used to control for multiple testing. For this BAYESCAN defines a q-value which represents the minimum FDR at which a locus may become significant. For this study a q-value of ≤ 0.1 was used.

Detection of directional selection

In order to determine the presence of candidate loci under directional selection associated with pollution in the Maipo River basin, populations located in non-polluted sites (NP) were compared with populations inhabiting polluted sites (P). To prevent the detection of outliers due to genetic drift or selection not related to pollution, all loci detected as outliers in the NP-NP comparisons were excluded. As a second control, candidate loci under directional selection which arose from the paired comparisons performed within the control basins Limarí, Choapa and Aconcagua River were also excluded from the list of candidate loci found in the NP-P comparison in the Maipo River. It is important to note that the three sites within each control basin represent different populations according to the F_{ST} paired analysis performed with the AFLP loci obtained in this study. With this analysis we aimed at avoiding candidate loci due to factors other than pollution.

Detection of balancing selection

To determine candidate loci under balancing selection related to pollution, a comparison of polluted populations (P-P) was performed. All loci detected as being under balancing selection in the NP-P or NP-NP comparisons within the Maipo basin or in the control basins (Limari, Choapa and Aconcagua) were excluded from the final analysis.

Discriminant Analysis of Principal Components (DAPC)

After detecting loci under selection related to pollution, three datasets of loci were constructed: i) candidate loci under directional selection in the PEL population and the three populations in non-polluted areas, ii) candidate loci under directional selection in the MEL population and the three populations in non-polluted areas; iii) candidate loci under balancing selection in the PEL and MEL populations. These three data sets were the input for the population structure analysis through a Discriminant Analysis of Principal Components (DAPC) implemented in the R software (R Core Team 2013). The ADE4 package allows for the performance of a multivariate analysis designed to identify numbers of clusters of genetically related individuals (Jombart et al. 2010).

RESULTS

In total, 149 AFLP loci were scored for all individuals. Because a number of loci were monomorphic in some sites, this number was lower in some comparisons. For example, in the Limarí, Choapa and Aconcagua River basins the number of polymorphic loci varied from 84 to 135, while comparisons in the Maipo River basin showed polymorphic loci ranged from 143 to 149 (Table 1).

Detection of selection

Directional selection

Using the DFDIST software, a total of thirteen loci were detected under directional selection in all comparisons. These comparisons were performed using a 95% confidence level and an FDR < 0.1. In this analysis, three loci constituted the maximum number of outliers per pair of comparisons. A total of 10 loci were identified, ranking from 1 to 3 loci (0.67 to 2.01% of the polymorphic loci of each comparison) in the 10 paired comparisons within the Maipo River basin. Comparisons performed within each control basin (Limarí, Choapa and Aconcagua) revealed the presence of 4 loci under selection: three detected in the Limarí and one in the Choapa basin (Table 1). After excluding all the loci detected under directional selection in the control basins and in the NP-NP comparisons, the final number of loci identified as candidates under directional selection in the NP-P comparisons was six (Table 2). In detail, between 0 and 2 (0 to

1.36% of the loci) were detected as candidates under directional selection in the paired NP-P comparisons, from the total of 6 loci identified, three related to the PEL population (loci 37, 55 and 72) and three to the MEL population (loci 103, 134 and 147). None were shared by both populations (Table 3). The result of one NP-P comparison performed with DFDIST implemented in the MCHEZA software is shown in Fig. 2.

Using the BAYESCAN software, a total of six loci were detected under directional selection in all comparisons, which were also identified by the DFDIST software. In the control basins, 3 loci (loci 111, 112 and 124) were under directional selection, ranking from 0 to 3 loci (0 to 2.9%) in each paired comparisons; two of these loci showed a "decisive" posterior probability > 0.99 and FDR < 0.01, while the third showed a "strong" posterior probability of 0.96 and FDR of 0.01. In the Maipo River basin, BAYESCAN software detected 3 loci (Table 1); locus 25 was identified in 6 out of 10 paired comparisons with a "strong" or "decisive" posterior probability, from 0.95 to 0.99 and FDR < 0.05. Locus 72 presented "strong" posterior probability of 0.95 and FDR of 0.05. Finally, locus 147 was detected in two of the comparisons with a "substantial" posterior probability of 0.85 and 0.90 and FDR < 0.08.

Excluding all the loci found under directional selection in the control basins (Limarí, Choapa and Aconcagua) and in the NP-NP comparisons (within Maipo basin), two loci (1.3% of the total loci) were retained identified as under selection in the NP-P comparison within the Maipo River. Locus 72 was detected in the comparison of PU (NP) with PEL (P), whereas locus 147 was detected in the comparison of MASFM (NP)

and MEL (P) (Table 3). These two loci were also detected for the same comparisons with the DFDIST software. The result of one NP-P comparison performed with BAYESCAN is shown in Fig. 3.

Thus combining the outliers found by both methods (i.e., DFDIST and BAYESCAN) a total of six loci (4% of the total loci) were identified as candidates to be under directional selection when NP and P populations were compared, two detected by both software programs and four detected only by the DFDIST software (Fig. 4a). Due to the diverse control methods applied to identify the loci under directional selection related to pollution, the four loci only identified by the DFDIST software were also considered reliable and were retained, along with the two loci detected by both methods, for the DAPC analysis. Of the six loci retained, three were under selection in the PEL population and 3 in the MEL population (Fig. 4b).

Table 1. Number and percent of candidate loci under directional selection identified with DFDIST and BAYESCAN software. The number of AFLP loci identified as candidates by both methods is also shown in the last column.

Basin	n	Number of polymorphic loci	Number and percent of candidate loci		Number of shared loci
			DFDIST	BAYESCAN	
Limarí					
CM-HU	60	84	2 (2.38%)	2 (2.38%)	2
CM-PA	60	103	3 (2.91%)	3 (2.91%)	3
HU-PA	60	103	0	0	0
Choapa					
CP-CA	59	128	1 (0.78%)	0	0
CP-CH	59	130	0	0	0
CA-CH	60	135	0	0	0
Aconcagua					
LL-PN	59	121	0	0	0
LL-PO	59	128	0	0	0
PN-PO	60	113	0	0	0
Maipo					
NP-P					
PU-PEL	47	146	1 (0.68%)	1 (0.68%)	1
IMPN-PEL	70	148	1 (0.68%)	1 (0.68%)	1
MASFM-PEL	71	149	2 (1.34%)	0	0
PU-MEL	46	148	2 (1.35%)	1 (0.68%)	1
IMPN-MEL	69	147	2 (1.36%)	0	0
MASFM-MEL	70	147	2 (1.36%)	2 (1.36%)	2
NP-NP					
PU-IMPN	71	149	3 (2.01)	1 (0.67%)	1
PU-SFMMA	72	149	1 (0.67%)	0	0
IMPN-SFMMA	95	149	1 (0.67%)	1 (0.67%)	1
P-P					
PEL-MEL	45	143	2 (1.4%)	2 (1.4%)	2

n: number of individuals; NP-P: non-polluted and polluted population comparison; NP-NP: comparison between two non-polluted populations; P-P: comparison between the two polluted populations.

Table 2. Loci under directional selection detected in the silverside fish inhabiting the Maipo River basin, obtained with DFDIST method. This table includes loci detected under directional selection in the other basins, the comparison NP-NP within the Maipo River basin and the loci retained for NP and P comparison after applying the control methods.

	Locus number
Loci detected under directional selection within Maipo River	11, 16, 25, 37, 55, 72, 103, 112, 134, 147
Loci detected under directional selection in control basin	95, 111, 112, 124
Loci detected in comparison NP *vs.* NP	11, 16, 25, 112
Loci retained under directional selection	37, 55, 72, 103, 134, 147

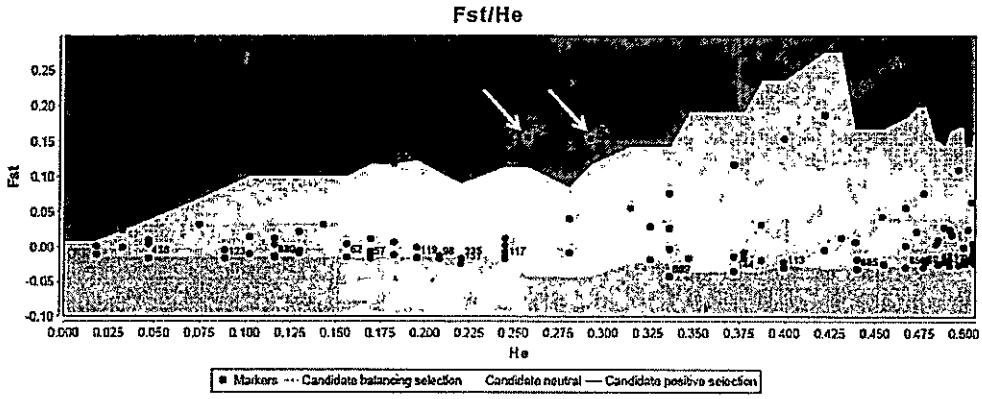


Figure 2 Example of an analysis performed with DFDIST implemented in MCHEZA. Distribution of F_{ST} values for each locus as a function of locus heterozygosity for the comparison of the IMPN – MEL (NP-P) population pair. Candidate loci identified by DFDIST are located in the dark gray (positive selection) and gray (balancing selection) regions with neutral loci in the middle region (light gray). Loci indicated by an arrow are those identified as under directional selection at 95% confidence level and with FDR < 0.1.

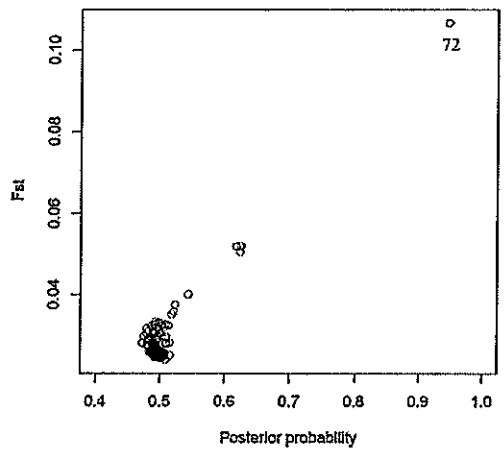


Figure 3 Example of an analysis performed with BAYESCAN: plot of F_{ST} values against posterior probability estimates for the comparison of the PU – PEL (NP-P) population pair. The labelled marker (locus 72) is a locus detected as candidate loci under directional with a posterior probability of 0.95 and FDR of 0.05.

Table 3. Number and percent of candidate loci (after the control methods) under directional selection for the comparison NP and P in the Maipo River basin. The analysis was performed with the software DFDIST and BAYESCAN.

Populations	n	Number of polymorphic loci	Number and percent of candidate loci pollution selection		Total loci under directional selection
			DFDIST	BAYESCAN	
NP-P					
PU-PEL	47	146	1 (0.68%)	1 (0.68%)	1 (0.68%)
IMPN-PEL	70	148	0	0	0
MASFM-PEL	71	149	2 (1.34%)	0	2 (1.34%)
PU-MEL	46	148	0	0	0
IMPN-MEL	69	147	2 (1.36%)	0	2 (1.36%)
MASFM-MEL	70	147	1 (0.68%)	1 (0.68)	1 (0.68%)
Total	166	149	6 (4.03%)	2 (1.34%)	6 (4.03%)

n: number of individuals; NP-P: non-polluted and polluted population comparison

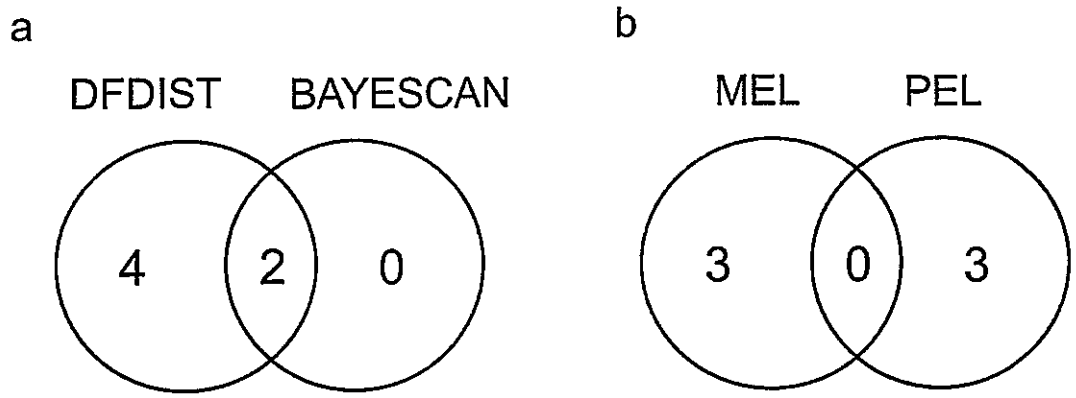


Figure 4. a) Venn diagram summarizing the number of loci detected under directional selection by one or both methods used, DFDIST and BAYESCAN; b) Venn diagram summarizing the number of loci detected as under directional selection related to pollution in one or both polluted populations (MEL and PEL).

Balancing selection

Detection of candidate loci under balancing selection showed different numbers of candidate loci per basin. The DFDIST software did not identify outlier loci for balancing selection in fish collected in the Limarí basin, whereas two of the three comparisons detected 18 loci outliers (14.1% of loci for all paired comparisons) in the Choapa River basin. Within the Aconcagua basin, six loci (5.3%) were detected as outliers in one comparison and eight loci (6.6%) in another comparison. For *B. microlepidotus* from the Maipo River basin, candidate loci under balancing selection were detected in all the comparisons performed, ranking from 10 to 34 loci (6.76-23.13% of the polymorphic loci of each paired comparison) (Table 4). Regarding the comparison of the two polluted populations (PEL-MEL), 24 loci (loci 8, 15, 20, 29, 30, 35, 46, 47, 57, 76, 79, 82, 83, 88, 94, 104, 114, 115, 119, 120, 125, 127, 132 and 149) were found as candidates for balancing selection (16.8% of polymorphic loci). After excluding loci found as outliers in the NP-P and NP-NP comparisons and those detected in the control basins, we considered four loci as candidates for being under balancing selection (loci 20, 79, 82 and 88).

The BAYESCAN software showed no evidence of balancing selection in the comparisons performed. In general BAYESCAN is more conservative in detecting loci under selection than the DFDIST software. The identification of loci under balancing selection related to pollution with the DFDIST method was considered reliable, due to the multiple control methods used and the fact that frequencies are relatively

homogeneous in the two polluted populations (Table 5). Overall this study showed that 10 loci (6.7% of the total loci) were identified as candidate loci under selection related to pollution, namely six (4.0%) as candidate loci under directional selection and four (2.7%) as loci under balancing selection.

Table 4. Number and percent of candidate loci under balancing selection identified for the comparisons performed in the four basins. Analysis performed with DFDIST software.

Basin	n	Number of polymorphic loci	Number and % of candidate loci (DFDIST)
Limarí			
CM-HU	60	84	0
CM-PA	60	103	0
HU-PA	60	103	0
Choapa			
CP-CA	59	128	18 (14.06%)
CP-CH	59	130	16 (12.31%)
CA-CH	60	135	0
Aconcagua			
LL-PN	59	121	8 (6.61%)
LL-PO	59	128	0
PN-PO	60	113	6 (5.31%)
Maipo			
NP-P			
PU-PEL	47	146	23 (15.75%)
IMPN-PEL	70	148	19 (12.84%)
MASFM-PEL	71	149	25 (16.68%)
PU-MEL	46	148	10 (6.76%)
IMPN-MEL	69	147	34 (23.13%)
MASFM-MEL	70	147	18 (12.24%)
NP-NP			
PU-IMPN	71	149	26 (17.45%)
PU-SFMMA	72	149	29 (19.46%)
IMPN-SFMMA	95	149	12 (8.05)
P-P			
PEL-MEL	45	143	24 (16.78%)

n: number of individuals; NP-P: non-polluted and polluted population comparison; NP-NP: comparison between two non-polluted populations; P-P: comparison between the two polluted populations.

Table 5. Frequency of the dominant allele for each of the four loci detected under balancing selection in *B. microlepidotus* obtained from the polluted site MEL and PEL

	MEL	PEL
Locus 20	0.57	0.59
Locus 79	0.57	0.59
Locus 82	0.52	0.54
Locus 88	0.7	0.65

Discriminant Analysis of Principal Components (DAPC)

The three AFLP sub-data sets showed that the first two principal components explained more than 79% of the total variance. For the case of candidate loci under directional selection, there was a clear segregation of PEL (P) from the three non-polluted populations (Fig. 5a). This segregation was not so clear in the case of the DAPC performed for the MEL population (P) (Fig. 5b), probably due to the few loci under directional selection related to pollution that MEL population has with two of the three populations of *B. microlepidotus* in the NP sites, namely MASFM and PU (Table 3). Finally, a clear superposition of the polluted populations was observed using the four loci detected under balancing selection (Fig. 5c).

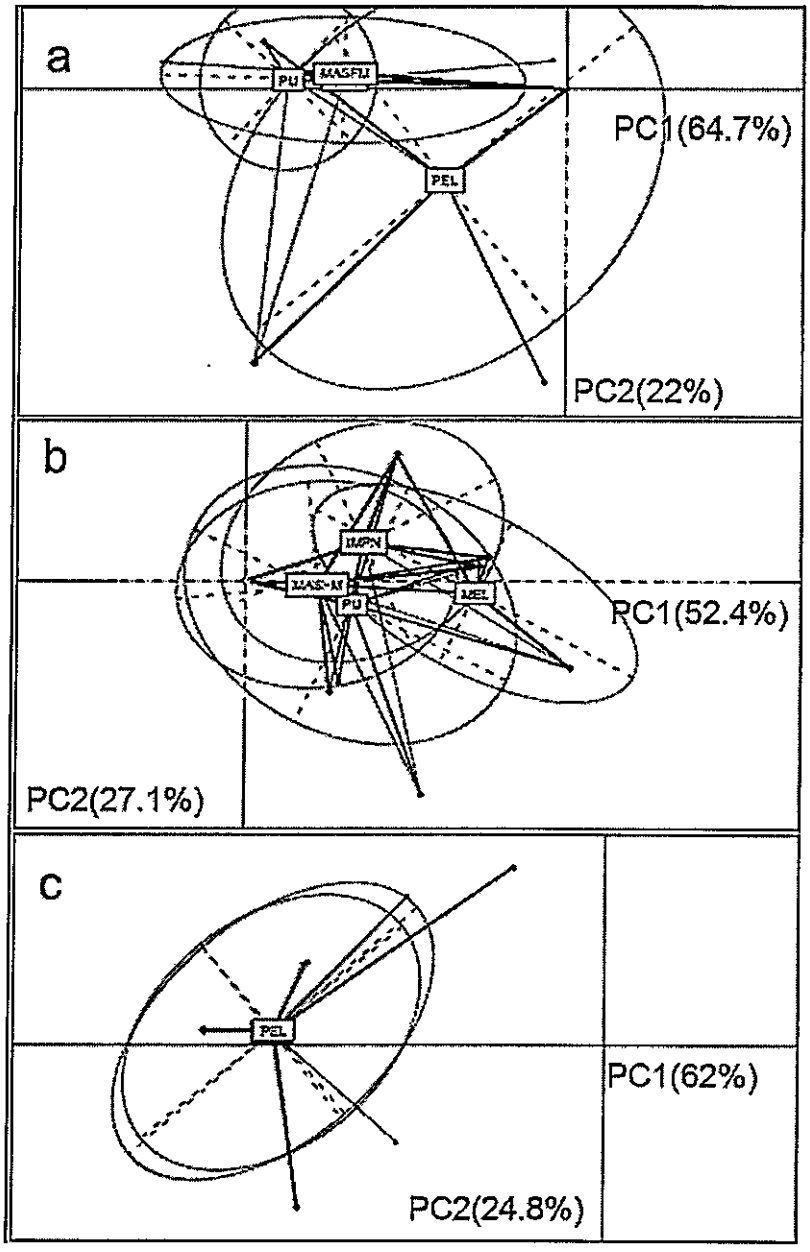


Figure 5. Discriminant Analysis of Principal Components for a) candidate loci under directional selection in PEL population, b) candidate loci under directional selection in MEL population and c) candidate loci under balancing selection in the PEL and MEL populations.

DISCUSSION

The analyses performed in this study showed the presence of selection in the *Basilichthys microlepidotus* genome within the Maipo River basin. Several loci in this species in this basin presented evidence of directional selection when populations inhabiting areas with pollution were compared to others without pollution. In addition, some evidence was observed for balancing selection between polluted populations.

Regarding directional selection, six loci (4% of the total loci analyzed) were identified under this type of selection when fish located in polluted sites were compared with those in non-polluted sites; three (2% of total loci) were detected through comparisons with each polluted site (PEL and MEL). This amount of directional selection agrees with other studies performed with different levels of pollution. For example, Paris et al. (2010) found that 3.2% of loci studied were under directional selection in the mosquito *Aedes rusticus* when treated with an insecticide, while Meyer et al. (2009) found that only 0.48% of the AFLP loci studied were under directional selection regarding the plant *Arabidopsis halleri* exposed to pollution. In the case of the freshwater fish *Perca flavescens* (Mitchill), Bélanger-Deschênes et al. (2013) showed that 6.3% of the single nucleotide polymorphisms (SNPs) were under directional selection when fish were exposed to heavy metals, while Williams and Oleksiak (2008) found between 1 and 6% of loci under directional selection for the marine fish *Fundulus heteroclitus* in sites with chronic pollution exposure. Interestingly, these authors did not find loci under selection sharing the three sites studied; only two sites shared loci. Thus most of the outlier loci

detected were unique to a single polluted population rather than shared across polluted populations, suggesting that different loci are involved in the process of adapting to a particular pollutant or stressing factor.

In our study, the analysis did not show evidence of the same loci under directional selection for populations of *B. microlepidotus* inhabiting polluted sites. This is not surprising, since the characteristics of the pollutants along the river are different. Vega-Retter et al. (2014) showed that the two polluted sites (MEL and PEL) differ in levels of pollution, with clear differences in dissolved oxygen as well as concentrations of copper and molybdenum. Thus it is probable that pollution affects natural populations in different ways, thus triggering adaptation to local characteristics of the pollution even within a basin.

The methods used in this study (DFDIST and BAYESCAN) detected different numbers of candidate loci; while the DFDIST software showed 6 loci (4%) under directional selection related to pollution, the BAYESCAN software identified 2 of these loci (1.3%) as candidates. Perez-Figueroa et al. (2010) compared three different methods, namely DFDIST, DETSELD and BAYESCAN, and concluded that the BAYESCAN software appears to be more efficient than the other methods in detecting loci under directional selection. However, Perez-Figueroa et al. (2010) also showed that DFDIST with a multi-test correction seems to be efficient with low values of neutral gene frequency differentiation ($F_{ST} \approx 0.025$). The MCHEZA software used in this study implements the DFDIST method with a multi-test correction, and the F_{ST} among our populations (after

excluding the loci identified under directional selection related to pollution) had low values that varied between 0.0022 and 0.07. Moreover, considering the differences between the methods used to detect loci under selection, to avoid type I error in the analysis we used two controls, comparisons between populations in the same basin and comparisons within basins with low pollution levels. With all these restrictions imposed on the data, the analyses revealed the presence of a minimum of 2 (BAYESCAN) and a maximum of 6 (DFDIST) loci under directional selection. All this background made us consider that the four loci detected only by the DFDIST software are reliable candidates loci under directional selection related to pollution. These results are also confirmed by the DAPC, where a segregation of MEL and PEL from the three non-polluted populations was clearly observed. It is important to note that this effect is clearer in the PEL population than in MEL, indicating that the pollution in the PEL site is probably more intense.

Studies searching for balancing selection have usually shown high percentages of detected loci. For example, Makinen et al. (2008) indicated that a high proportion of loci (14.7%) might be affected by balancing selection between marine and freshwater three-spined stickleback (*Gasterosteus aculeatus* Linnaeus), while Akey et al. (2002) found that 11% of the total SNPs studied were under balancing selection when the analysis was performed in three human populations. On the other hand, Bélanger-Deschênes et al. (2013) detected that 16.6% of the SNPs studied were under balancing selection comparing the fish *Perca flavescens* exposed and not exposed to metals. Despite all this information we do not know of studies that have been performed to detect balancing

selection related to pollution in nature. In our case, four loci (2.7% of total loci) were identified as loci candidates under balancing selection when the comparison between the populations inhabiting polluted areas was performed.

DFDIST detected four loci as candidates under balancing selection related to pollution, while BAYESCAN did not show evidence of balancing selection in any of the comparisons performed. Despite these clear differences between the two methods, we considered that the four loci detected by DFDIST were reliable due to the different restriction imposed to the analysis: i) by using three control methods in order to exclude loci detected under balancing selection in the NP-P or NP-NP comparisons within the Maipo basin as well as those detected in the control basins; ii) the fact that the four loci detected were observed to have similar frequencies in both polluted populations; and iii) the remarkable similarity of these populations in the DAPC based on these four loci.

Considering that balancing selection could maintain alternatively adaptive alleles in environments where the selective pressures change in space or time (Levene 1953; Hedrick et al. 1976) our study pointed out pollution as a force that could maintain polymorphism. In our case, Mediterranean rivers in central Chile impose spatial and temporal heterogeneity, thus making it possible to observe the action of balancing selection as found in the silverside *B. microlepidotus* within the Maipo River basin. This river presents seasonal variations related to changes in temperature, rainfall and the gradual melting of snow accumulated in the mountain range (Dirección general de Aguas 2004d) and inter-annual variations mainly related to the El Niño Southern

Oscillation (ENSO) climatological phenomenon (Rutllant and Fuenzalida 1991). Consequently, with these changes in flow, variations in the chemical composition of the water and sediment are also expected, this being a possible explanation for the loci under balancing selection related to pollution in this basin.

This study contributes to the understanding of the selective effects that pollution may have, and shows that selective pressures could be different even within the same basin. To our knowledge this is the first study that identifies loci under balancing selection related to pollution, hence indicating that pollution could be responsible for maintaining polymorphism in the selected loci. Finally, by viewing these results together and based on the observation that gene diversity and migration rates for this species do not seem to be affected by pollution (Vega-Retter et al. 2014), we consider that this is a good study model to understand better the mechanisms involved in adapting to pollution. The next step in our work will be the detection of genes with differential expression between polluted and non-polluted populations.

Acknowledgements

We thank C Quezada-Romegialli, JP Oyanedel and P Muñoz-Rojas for helping us in the field work. Thanks to JP Oyanedel for comments on an earlier version. DV thanks the financial support of Fondecyt 1100341, Proyecto Enlace Universidad de Chile 2013, Basal Grant PFB 023 and ICM P05-002; CVR thanks Conicyt Doctoral Fellowship 21090188 and doctoral thesis fellowship 24121005.

REFERENCES

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. Genome Research 12: 1805–1814.

Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Research 25: 4692–4693.

Antao T, Beaumont MA (2011) Mcheza: A workbench to detect selection using dominant markers. Bioinformatics 27: 1717–1718.

Bahamondes I, Soto D, Vila I (1979) Hábitos alimentarios de las especies de Atherinidae del Embalse Rapel. Medio Ambiente 4: 3-18.

Beaumont MA. Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society of London series B 263:1619–1626.

Beaumont MA (2005) Adaptation and speciation: what can FST tell us?. Trends in Ecology & Evolution 20: 435–440.

Bélanger-Deschênes S, Couture P, Campbell PGC, Bernatchez L (2013) Evolutionary change driven by metal exposure as revealed by coding SNP genome scan in wild yellow perch (*Perca flavescens*). Ecotoxicology 22: 938-957.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative Genome Scan to Detect Candidate Loci for Adaptation Along a Gradient of Altitude in the Common Frog (*Rana temporaria*). Molecular Biology and Evolution 23: 773-783.

Caballero A, Quesada H, Rol´an-Alvarez E (2008). Impact of amplified fragment length polymorphism size homoplasy on the estimation of population genetic diversity and the detection of selective loci. Genetics 179:539–554.

Campbell D, Bernatchez L (2004) Genetic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. Molecular Biology and Evolution 21: 945-956.

Campos H, Dazarola G, Dyer B, Fuentes L, Gavilan J, Huaquin L, Martinez G, Melendez R, Pequeño G, Ponce F, Ruiz V, Siefeld W, Soto D, Vega R, Vila I (1998) Categorías de conservación de peces nativos de aguas continentales de Chile. Boletín del Museo Nacional de Historia Natural 47: 101-122.

Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. Plos Genetics 2: 379–384.

Comte S, Vila I (1992) Spawning of *Basilichthys microlepidotus* (Jenyns). Journal of Fish Biology 41: 971-981.

Dirección General de Aguas (2004a) Diagnóstico y clasificación de los cursos y cuerpos de agua según objetivos de calidad, cuenca del río Aconcagua. Ministerio de Obras Públicas, Gobierno de Chile.

Dirección General de Aguas (2004b) Diagnóstico y clasificación de los cursos y cuerpos de agua según objetivos de calidad, cuenca del río Choapa. Ministerio de Obras Públicas, Gobierno de Chile.

Dirección General de Aguas (2004c) Diagnóstico y clasificación de los cursos y cuerpos de agua según objetivos de calidad, cuenca del río Limarí. Ministerio de Obras Públicas, Gobierno de Chile.

Dirección General de Aguas (2004d) Diagnóstico y clasificación de los cursos y cuerpos de agua según objetivos de calidad, cuenca del río Maipo. Ministerio de Obras Públicas, Gobierno de Chile.

Duarte W, Feito R, Jara C, Moreno C, Orellana AE (1971) Ictiofauna del sistema hidrográfico del río Maipo. Boletín del Museo Nacional de Historia Natural Chile 32: 227-268.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180: 977–993.

Freeland JR, Biss P, Conrad KF, Silvertown J (2010) Selection pressures have caused genome-wide population differentiation of *Anthoxanthum odoratum* despite the potential for high gene flow. Journal of Evolutionary Biology 23: 776-782.

Hedrick PW, Ginevan ME, Ewing EP (1976) Genetic polymorphism in heterogeneous environments. Annual Review of Ecology and Systematics 7: 1–32.

Henry P, Russello A (2013) Adaptive divergence along environmental gradients in a climate-change-sensitive mammal. Ecology and Evolution 3: 3906-2917.

Hoffmann AA, Willi Y (2008) Detecting genetic responses to environmental change. Nature Reviews Genetics 9: 421–432.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC genetics 11: 94.

Levene H (1953) Genetic equilibrium when more than one ecological niche is available. American Naturalist 87: 331-333.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003). The power and promise of population genomics: from genotyping to genome typing. Nature Reviews Genetics 4: 981–994.

Makinen HS, Cano JM, Merilä J (2008) Identifying footprints of directional and balancing selection in marine and freshwater three-spined stickleback (*Gasterosteus aculeatus*) populations. Molecular Ecology 17: 3565-3582.

Meyer CL, Vitalis R, Saumitou-Laprade P, Castric V (2009) Genomic pattern of adaptive divergence in *Arabidopsis halleri*, a model species for tolerance to heavy metal. Molecular Ecology 18: 2050-2062.

Meudt HM, Clarke AC (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances. Trends in Plant Sciences 12:106–117.

Muñoz BM (2007) Cambios en la riqueza íctica del rio Maipo. Tesis Biología con Mención en Medio Ambiente, Facultad de Ciencias, Universidad de Chile.

Nielsen R (2005) Molecular signatures of natural selection. Annual Reviews of Genetics 39: 197–218.

Nielsen EE, Hemmer–Hansen J, Foged-Larsen P, Bekkevold D (2009) Population genomics of marine fishes: identifying adaptive variation in space and time. Molecular Ecology 18: 3128–3150.

Pardo R, Vila I, Bustamante R (2008) Demographic response of *Stratiodrilus aeglaphilus* (Anelida, Histriobdellidae) to organic enrichment: experimental assessment. Hydrobiologia 598: 11-16.

Paris M, Boyer S, Bonin A, Collado A, David JP, Despres L (2010) Genome scan in the mosquito *Aedes rusticus*: population structure and detection after insecticide treatment. Molecular Ecology 19: 325-337.

Pérez-Figueroa A, García-Pereira MJ, Saura M, Rolán-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. Journal of Evolutionary Biology 23: 2267–2276.

Quezada-Romegialli C, Fuentes M, Veliz D (2010) Comparative population genetics of *Basilichthys microlepidotus* (Atheriniformes: Atherinopsidae) and *Trichomycterus areolatus* (Siluriformes: Trichomycteridae) in north central Chile. Environ Biol Fish 89:173–186.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Reusch TBH, Wood TE (2007) Molecular ecology of global change. Molecular Ecology 16: 3973–3992.

Rutllant J, Fuenzalida H (1991) Synoptic aspects of central rainfall variability associated with the Southern Oscillation. *International Journal of Climatology* 11: 63–76.

Schlötterer C (2003) Hitchhiking mapping–functional genomics from the population genetics perspective. Trends in Genetics 19: 32–38.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. Molecular Ecology 14: 671–688.

Vega-Retter C, Muñoz-Rojas P, Vila I, Copaja S, Véliz D (2014) Genetic effects of living in a highly polluted environment: the case of the silverside *Basilichthys microlepidotus* (Jenyns) (Teleostei: Atherinopsidae) in the Maipo River basin, central Chile. Population Ecology: DOI 10.1007/s10144-014-0444-3.

Véliz D, Duchesne P, Bourget E, Bernatchez L (2006) Stable genetic polymorphism in heterogenous environments: balance between asymmetrical dispersal and selection in the acorn barnacle. Journal of Evolutionary Biology 19: 589-599.

Véliz D, Catalán L, Pardo R, Acuña P, Diaz A, Poulin E, Vila I (2012) The genus *Basilichthys* (Teleostei: Atherinopsidae) revisited along its Chilean distribution range (21° to 40° S) using variation in morphology and mtDNA. Revista Chilena de Historia Natural 85: 49-59.

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Friters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23: 4407–4414.

Williams LM, Oleksiak MF (2008) Signatures of selection in natural populations adapted to chronic pollution. BMC Evolutionary Biology 8: 282-293.

World Water Assessment Programme (2009) The United Nations World Water Development Report 3: Water in a Changing World. UNESCO, Paris and Earthscan, London.

CHAPTER 3

LIVING IN A HIGHLY POLLUTED ENVIRONMENT: CHANGES IN GENE EXPRESSION AND CANDIDATE SNPS IN THE SILVERSIDE *BASILICHTHYS MICROLEPIDOTUS* (JENYNS) (TELEOSTEI: ATHERINOPSIDAE) INHABITING THE MAIPO RIVER BASIN, CENTRAL CHILE

Article in preparation

ABSTRACT

Understanding the molecular basis of the biological adaptation to environmental changes has become an important goal in environmental biology especially with the increasing habitat degradation as a consequence of human activities, being pollution one of the major forces affecting freshwater systems in the world. The development of the massive parallel next generation sequencing technologies has made possible the study of the genetic adaptation for all types of organisms, making possible the studies of gene expression in order to determine the transcriptomic and adaptive responses to environmental changes such as pollution. In Chile, the Maipo River basin is highly polluted and has experienced water quality deterioration and eutrophication, mostly as a product of organic matter from untreated sewage. In this basin, populations of the silverside *Basilichthys microlepidotus* inhabit polluted and non-polluted areas, showing evidence that 4% of AFLP loci were under directional selection between these fish. Thus, the objective of this study was to identify genes with differential expression (using RNA-seq technique) between individuals collected from populations inhabiting non-polluted and polluted areas and to determine differences in single nucleotide polymorphisms (SNP) between these populations. After an analysis that imposed several restrictions to avoid false positive results, a total of six genes with differential expression were identified, two down-regulated and four up-regulated in the individuals from polluted areas. Also, two candidates SNPs were detected when individuals sampled in polluted and non-polluted populations were compared; both SNPs were located in mitochondrial genes. These findings, along with the identification of genes whose

expression profiles were modified, provide new insights on our understanding on the species adaptation to the changing environments including those linked with pollution.

Key words: gene expression, single nucleotide polymorphisms, pollution.

INTRODUCTION

Understanding the molecular basis of the biological adaptation to environmental changes has become an important goal in environmental biology. The ability of populations to adapt to these changes is one of the main focuses of the emerging field of ecological genomics (Ungerer et al. 2008), and this fact becomes important with the increasing habitat degradation as a consequence of human activities.

Until recently, genome-wide studies were unattainable in the non-model organisms; nevertheless the development of massive parallel next generation sequencing technologies (NGS) makes possible the study of genetic adaptation in this group of organisms. In this context, NGS has allowed to perform studies of gene expression in all species in order to determine the transcriptomic response to environmental changes (Ekblom and Galindo 2011). These types of studies represent a successful way to assess genes involved in adaptation, as it is known that transcription levels have great potential for evolutionary novelty, being the variation in gene expression an important mechanism for evolution by natural selection (Oleksiak et al. 2002; Fay and Wittkopp 2008). Then, the expression of functional genes plays an important role in physiological acclimation, but it is also important in the evolutionary adaptation to changes in the environment or in the use of new habitats.

In the last century, pollution due to human activities has been one of the major force affecting freshwater systems in the world (World Water Assessment Programme

2009).In this context, new studies focus in the detection of changes in gene expression in fishes exposed to diverse pollutants. For example, Garcia et al. (2012) indicated that there is a complex genomic response in the freshwater fish *Fundulus grandis* (Baird and Girard) exposed to oil, with 1,070 down-regulated and 1,251 up-regulated genes, while Huang et al. (2012) found differential gene expression related to neurobehavioral defects, mitochondrial dysfunction and the metabolism of proteins and fats when the transcriptional response of *Oryzias melastigma* (McClelland) embryos were investigated after exposure to the organic pollutant Perfluorooctane sulfonate. On the other hand, Olsvik et al. (2013) analyzed the gene expression in the liver of the freshwater fish *Lota lota* (Linnaeus) from Lake Mjøsa (contaminated with organic pollutants) and compared it with individuals from the unpolluted Lake Losna. The gene expression profile suggested enrichment in the mechanism associated with drug metabolism and in the oxidative stress of organisms living in the Lake Mjøsa.

Pollution not only produces change in gene expression, SNPs were also found related with the metal contamination. Belanger-Deschenes et al. (2013) determined functional polymorphisms in chronically metal-contaminated wild yellow perch (*Perca flavescens*, Mitchill); they found three SNPs that putatively could be under selection. Moreover, the allelic frequencies of these three SNPs were mostly correlated with the population mean concentration of hepatic cadmium. Then, the authors proposed that Cd represents a selective agent driving an evolutionary change in the population of *P. flavescens* inhabiting contaminated areas.

In Chile, the Maipo River is one of most polluted basin that receives the effect of mining from the Andes Mountains and effluents from factories surrounding the basin (Dirección General de Aguas 2004). The Maipo River supplies water to 6.7 million inhabitants according to the most recent census in 2012 (40% of the Chilean population), but also receives the wastewater from these inhabitants. This basin has experienced water quality deterioration and eutrophication (Pardo et al. 2008), mostly as a product of organic matter from untreated sewage.

Studies on the fish diversity have shown a significant reduction in the richness and species abundance during the last 30 years in the Maipo basin (Muñoz 2007), probably associated with the habitat degradation and pollution. One species that has shown an important decrease in its abundance is the silverside *Basilichthys microlepidotus* (Jenyns), an atherinopsid endemic to Chile considered to be endangered (Campos et al. 1998). This species inhabits lakes and rivers from 28° to 39°S (Véliz et al. 2012) and feeds on insect larvae, small invertebrates, filamentous algae and detritus (Duarte et al. 1971; Bahamondes et al. 1987). Comte and Vila (1992) described that reproduction occurs from August to January. A previous study performed in the Maipo basin showed that populations of *B. microlepidotus* inhabit polluted and non-polluted areas of the river, specifically there are two populations inhabiting in polluted sites and three populations living in non-polluted areas. Moreover, they also point out that the differences between the non-polluted and polluted areas are mainly explained by eight out of seventeen measured environmental variables namely total dissolved solids, conductivity, chloride, sulfate, molybdenum, ammonium, copper and nitrite (Vega-

Retter et al. 2014). Further, Vega-Retter et al. (submitted) detected that 4% of the AFLP loci studied in *B. microlepidotus* were under directional selection when comparing fish inhabiting polluted and non-polluted areas. Thus, this system allows for the study of pollution effects in fish, in order to determine the responses of the organisms in the presence of diverse pollutants in their habitat. Consequently, *B. microlepidotus* and the Maipo basin arise as an interesting model system to study the physiological effects of the pollution and the molecular basis of pollution adaptation using a transcriptional approach (RNA-seq), which was performed in liver samples of *B. microlepidotus* inhabiting polluted and non-polluted areas, this because it is known that liver performs a number of important and complex biological functions that are essential for survival like nutrient synthesis, transformation and storage, as well as endogenous and exogenous substance detoxification (Hasenfuss et al. 2014). Therefore, the objective of this study was to identify genes with differential expression by comparing individuals collected from populations inhabiting non-polluted and polluted areas as well as to determine differences in single nucleotide polymorphism variability between these populations.

MATERIALS AND METHODS

Sampling sites and collecting samples

Four out of seven sampling sites from Vega-Retter et al. (2014) were chosen for this study; two sites were defined as non-polluted: San Francisco de Mostazal (SFM) and Isla de Maipo (IM), and two sites defined as polluted: Melipilla (MEL) and Pelvin (PEL) (Fig.1); each one of these sites represents an independent fish population. In November 2012, three individuals from each non-polluted site (mean length = 8.95 cm, SD = 3.17) and three from each polluted site (mean length = 6.72 cm, SD = 2.11) were collected, the liver tissues were immediately obtained and transported in RNA-later (Life Technologies) to the laboratory; after that, tissues were stored at -80°C until the RNA extraction.

Physicochemical differences among the collecting sites

A Principal Component Analysis (PCA) was performed in order to determine the variables that most contributed to the total variance and explain the differences among the four sites at the moment that the fish samples for the RNA-seq analysis were collected. By using the same variables measured by Vega-Retter et al. (2014), namely: conductivity (EC), pH, total dissolved solids (TDS), sulfate (SO^{2-}_4), nitrite (NO^-_2), chloride (Cl^-), ammonium (NH^+_4), nitrate (NO^-_3), phosphate (PO^{3-}_4), sodium (Na^+), potassium (K^+), calcium (Ca^{2+}) and magnesium (Mg^{2+}), dissolved oxygen concentration

(DO) and Biological Oxygen Demand (BOD_5). Metals were excluded in this analysis. For details of the sampling method and laboratory measurements please refer to Vega-Retter et al. (2014). In the PCA, the factor loadings were used to determine the contribution of the variables in each component. PCA analysis was performed using the ADE4 library implemented in the R software (R Core Team 2013).

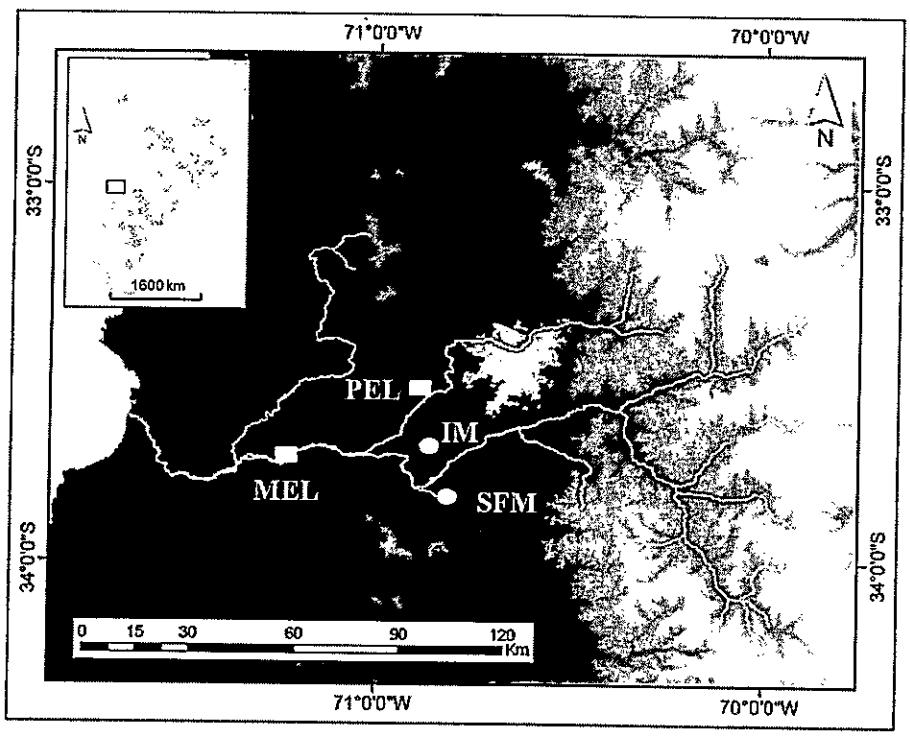


Figure 1. Samplig sites of *B. microlepidotus* in the Maipo River basin. Circles represent non-polluted sites and rectangles represent polluted sites according to Vega-Retter et al. (2014).

RNA extraction and sequencing

Total RNA from liver samples was extracted using the PureLink™ RNA Mini Kit (Ambion) following company instructions; quality and quantity of the Total RNA was checked using an Agilent Model 2100 Bioanalyzer at OMICS Solutions (Santiago, Chile). Samples were stored in RNase free water supplemented with Superase-In™ RNase Inhibitor (Ambion, Austin, TX, USA) at -80°C. Total RNA were subsequently enriched for polyA mRNA using the MicroPoly(A)Purist™ kit (Ambion) protocol. The mRNA from 11 samples was used to prepare 11 separate barcoded libraries with the Ion Total RNA-Seq Kit v2 (Life Technologies). An individual from the site SFM was excluded in this analysis. These 11 samples were sequenced in a total of four runs, three in the Ion Torrent platform using an Ion 318 chip, the fourth on the Ion Proton with the PI chip. Library construction and sequencing were performed by OMICS Solutions (Santiago, Chile).

Short read, quality control and de novo assembly

Short read and quality filtration was performed with the PRINSEQ (Schmieder and Edwards, 2011) and TRIMMOMATIC (Lohse et al. 2012) software. TRIMMOMATIC was used with a sliding window trimming of size 10 and a threshold of average quality of 15, also reads shorter than 60 bp were removed. PRINSEQ was used to remove reads with low mean phred scores (<Q13), the PolyA and PolyT tails at the 5' and 3' ends, reads shorter than 60 bp and to trim reads longer than 250 bp until this length, and also

the redundant reads were removed for the set of reads used to perform the *de novo* assembly .This assembly was performed using the MIRA assembler (Cheveruex et al. 1999), which is based on overlap graph algorithm. The overlap between each pair of reads is computed and compiled into a graph, where each node represents one short read, and an edge between two nodes indicates that the two short reads have overlapping sequences. After some steps of simplifying the overlap graph by removing the transitive nodes and edges, a chain of nodes elicits the sequence of a contig. For our analysis, the minimum number of reads per contig was set to 10. In order to create a set of non-redundant contigs, CD-HIT software was used (http://weizhong-lab.ucsd.edu/cd-hit/) with a similarity threshold of 90%.

Mapping of sequence reads, differential expression analysis and single nucleotide polymorphisms detection

To test for differential gene expression, individual reads for each *B. microlepidotus* were mapped back to the assembled transcriptome by using the alignment program TMAP (http://github.com/iontorrent/TMAP/tarball/tmap.0.3.7). The number of reads aligned to each contig for each sample was estimated with the IdxStats command of the SAMtools software (Li et al. 2009). An in-house Perl script was used to choose contigs with expression in at least 4 out of 5 individuals of the non-polluted condition and in least 5 out of 6 individuals of the polluted condition. The count data of this contigs was used as the input for the DESeq analysis (Anders and Huber, 2010) implemented in the MultiExperiment Viewer software (http://www.tm4.org/mev.html). DESeq estimates

variance-mean dependence in the data and tests for differential gene expression based on the negative binomial distribution. The five and six samples from each treatment (non-polluted and polluted respectively) were used to generate mean expression levels with associated variances. Differential expression was tested with an $\alpha = 0.05$ adjusting a match of 5% false discovery rate (FDR). In order to avoid contigs with differential expression due to one individual over expressed, we discarded the contigs that contained outliers with 1.5% more than the Interquartile Range (IQR) by regarding the boxplot in the R software (R Core Team 2013).

In order to detect single nucleotide polymorphisms (SNPs) between individuals of *B. microlepidotus* inhabiting non-polluted and polluted sites, the variant calling of SAMtools and an in-house Perl script were used. We considered as a valid SNP when the position had at least 5 reads per individual. Three different types of SNPs were considered as candidates to be related to the pollution:

i) Fixed SNPs. Those that presented a fixed base in fishes inhabiting polluted areas that was different from the fixed base in the individuals sampled in the non-polluted areas. For the bioinformatics analyses, the script considered one fixed base in one condition when this base was present in at least 60% of the total reads of four or more individuals of each condition,

ii) SNPs with a fixed base at the non-polluted condition but with diverse nucleotides at the same position in the polluted condition (FNP-DP). For the bioinformatics analyses,

those that showed the same nucleotide in a proportion higher than 85% in the total reads of the individuals of the non-polluted condition and not presenting one of the four bases in more than 40% of the total reads of the individuals inhabiting polluted sites were considered,

iii) SNPs with a fixed base at the polluted condition, but with nucleotide diversity at the non-polluted condition, FP-DNP. For the bioinformatics analyses, those that showed a fixation of a nucleotide higher than 85% in the total reads of the individuals of the polluted condition and not presenting one of the four bases in more than 40% of the total reads of the individuals inhabiting non-polluted sites.

The transcripts identified with differential expression by DESeq and the transcripts detected with SNPs potentially related to the pollution were blasted using blastx function of the Blast2GO software (Conesa et al. 2005). Blastx function was performed with a minimum E-value score of 1.0E-06. To allocate gene ontology (GO) terms to each annotated sequence, successful blast hits were mapped and annotated using Blast2GO with an annotation cut-off threshold set to 55 and the GO level weighting set to 5 as recommended by the authors of the software. The basic annotation obtained by retrieving GO terms from blastx matches was enriched and refined using InterProScan, both implemented in the Blast2GO software.

RESULTS

Physicochemical differences among the collecting sites

The PCA analysis showed that the first two components accounted for 87% of the total variance (Fig. 2), where the Principal Component 1 (60% of the total variance) had high loadings for pH, Na, Cl^- and TDS, and the Principal Component 2 (27% of the variance) had high loadings for NO^-_3, PO^{3-}_4, NH^+_4, and NO^-_2. It is important to note that four (Cl^-, TDS, NH^+_4, and NO^-_2) of the eight variables that accounted for most of the variance in this PCA were also observed by Vega-Retter et al. (2014). Moreover, the PCA showed a spatial separation of the polluted and non-polluted populations in a two dimension representation (Fig. 2).

Short read and quality filtration and de novo assembly

A total of 30.15 million reads (4.12 Gb) were obtained from the four runs performed. After of the trimming process, a total of 22.80 million reads (3.16 Gb) was retained for the mapping, with a final number of reads per individual ranged from 1.3 million (0.17 Gb) to 3.85 million (0.59 Gb). For the *de novo* assembly, 14.46 million reads (2.13 Gb) were used, 34,786 contigs longer than 200 bp were obtained, with the largest contig containing 6,472 bp and a N50 = 674. A total of 34,385 of non-redundant contigs were finally obtained.

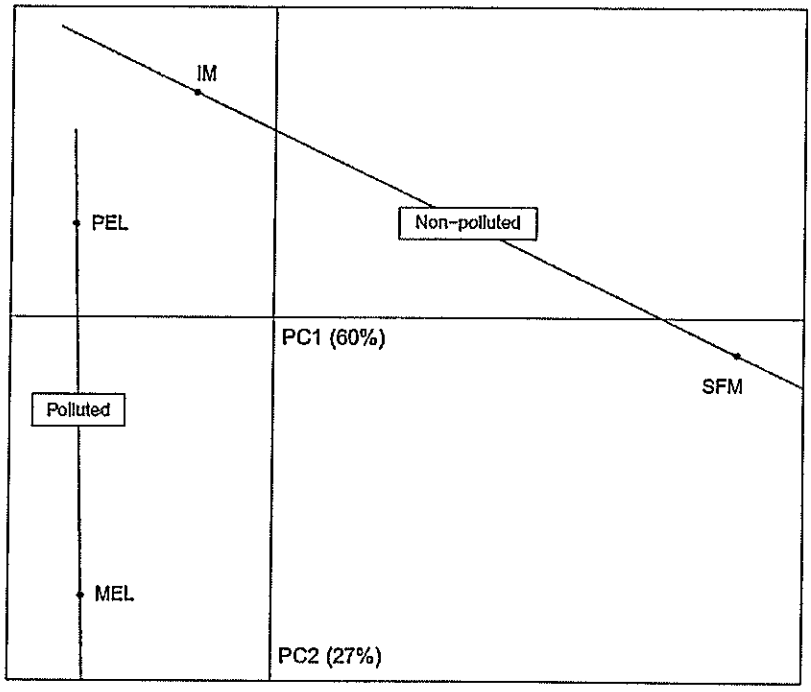


Figure 2. Principal Component Analysis performed with 15 environmental variables measured in the four sampling sites of *B. microlepidotus*. The non-polluted group contains the following sites: SFM, IM; the polluted group contains the MEL and PEL sites. PC1 and PC2 stand for Principal Components 1 and 2, respectively.

Mapping of sequence reads, analysis of differential expression and single nucleotide polymorphisms detection

The percentages of mapped reads were ranged from 61.27% to 87.25% for individual; the average was 74.23% with a standard deviation of 7.72. Contigs that showed expression in the most of individuals were in total 13,252. From these contigs, the DESeq software detected a total of 31 contigs with differential expression between the individuals inhabiting polluted areas and those inhabiting non-polluted sites. After

removing contigs with false differential expression (checking for outlier expression), a total of 6 contigs with differential expression were retained; from these 4 were up-regulated and 2 down-regulated in the polluted conditions regarding to the non-polluted condition.

In the SNPs detection, a total of 2 SNPs were identified: one fixed SNP and one FP-DNP SNP. In the case of the fixed SNP, an Adenine was observed in the position 1143 of the contig c11 in the fishes located in the non-polluted areas, while a Guanine was observed in the individuals of polluted areas. For the FP-DNP SNP, the Timine was fixed at the position 192 of the contig c138 in the polluted condition, while in the same position the four nucleotide bases were observed in the individuals from the non-polluted condition. More information in Table 1. The sequence description and the gene ontology of the contigs detected with differential expression by DESeq and with SNPs are shown in Table 2 and overexpression ratio at one of both conditions is shown in table 3. In the case of the contigs that present SNPs, these were involved in biological process as ATP synthesis coupled electron transport and aerobic respiration; for the contig that contains the FP-DNP SNP (sequence description: Cytochrome c oxidase subunit 1) the molecular function of iron binding was also detected. In the case of the contigs up-regulated in the pollution condition, genes were found involved in biological processes such as response to viruses, positive regulation of cell proliferation or regulation of transcription.

Table 1. Types of SNPs detected and its features.

Contig	Position	SNP type	Base fixed in non-polluted	Base fixed in polluted
c11	1143	Fixed	A	G
c138	192	FP-DNP	none (A = 35.3%; G = 23.53%, C = 22.05%, T = 19.12)	T (89.31%)

Table 2. Sequence description and gene ontology of genes with SNPs or differential expression between individuals from polluted and non-polluted sites.

				Gene ontology		
Sequence name	Sequence description	Sequence present	min. e value	Biological process	Molecular Function	Cellular component
c11	Nadh dehydrogenase subunit 5 (ND5)	Fixed SNP	0	ATP synthesis coupled electron transport	NADH dehydrogenase (ubiquinone) activity	Mitochondrial inner membrane; respiratory chain; integral to membrane
c138	Cytochrome c oxidase subunit 1 (COI)	FP-DNP	2.2 E-25	Aerobic respiration; oxidative phosphorylation; electron transport chain	Electron carrier activity; iron ion binding; heme binding; cytochrome-c oxidase activity	Mitochondrial inner membrane; respiratory chain; integral to membrane.
c2314	Ornithine decarboxylase-like (ODC-like)	Upregulation in P	0	Response to virus; eye photoreceptor cell development; kidney development; positive regulation of cell proliferation; putrescine biosynthetic process from ornithine.	Orithine decarboxylase activity	Cytosol

c2834	Phosphoenolpyruvate cytosolic	Upregulation in P	3 E-23	Gluconeogenesis; phosphorylation; response to glucose stimulus	GTP binding; kinase activity; phosphoenolpyruvate carboxykinase (GTP) activity	-
c3464	Cysteine serine-rich nuclear protein 1-like (CSRNP-1-like)	Upregulation in P	3.3 E-177	midbrain development; diencephalon development	-	-
c4323	Transcription factor Junb-like	Upregulation in P	6.3 E-167	Regulation of transcription, DNA-dependent; fin regeneration.	Sequence-specific DNA binding; sequence-specific DNA binding transcription factor activity	Nucleus
c5558	Probable E3 ubiquitin-protein ligase DTX2-like	Downregulation in P	1.1 E-109	-	Zinc ion binding	-
c18433	Chymotrypsin b-like	Downregulation in P	6.6 E-10	-	Serine-type peptidase activity	-

Table 3. Differential expression ratio for the overexpressed genes in one of both conditions.

Sequence name	Sequence description	Up-regulated in	Overexpression ratio
c2314	Ornithine decarboxylase-like (ODC-like)	Polluted condition	12.42
c2834	Phosphoenolpyruvate cytosolic	Polluted condition	9.32
c3464	Cysteine serine-rich nuclear protein 1-like (CSRNP-1-like)	Polluted condition	13.49
c4323	Transcription factor jun-b-like	Polluted condition	12.16
c5558	Probable E3 ubiquitin-protein ligase DTX2-like	Non-polluted condition	9.92
c18433	Chymotrypsin b-like	Non-polluted condition	6.61

DISCUSSION

The main goal of this study was to determine the effect of the pollution on the gene expression and SNP variability in the populations of *B. microlepidotus* inhabiting polluted areas in the Maipo River basin. In the PCA analysis performed with the fifteen environmental variables measured, a clear segregation between the polluted and non-polluted sites was observed. In total, 6 genes with differential expression were identified, 4 up-regulated in the individuals from the polluted areas and 2 down-regulated in these individuals. Furthermore, two candidate SNPs possibly related with the pollution were detected. In the case of the up-regulated genes, these encode for the following functions:

i) Ornithine decarboxylase-like (ODC-like). Ornithine decarboxylase is the first and rate limiting enzyme in polyamine biosynthesis, catalyzing the decarboxylation of L-ornithine to putrescine. The polyamine biosynthetic pathway is the critical regulator of cell growth, differentiation, and cell death, and the polyamines are involved in nucleic acid packaging, DNA replication, apoptosis, transcription, and translation (Jhingran et al., 2008). Thus, the ODC is indirectly related with the DNA replication, translation and also with its transcription. Whereas Sotomayor et al. (2012) and Yang et al. (2013) point out a decreasing activity in specimens of the frog *Rhinella arenarum* (Hensel) and the plant *Hydrocharis dubia* (Blume) exposed to organophosphorus pesticides and cadmium, Xu et al. (2011) and Qiao et al. (2012) showed an increase in its activity in the aquatic plants *Potamogeton crispus* (Linnaeus) and *Spirodela polyrhiza* (Linnaeus) exposed to lead. ODC activation and the increment of the polyamine concentration were

associated with tumor promotion and progression. It has also been supposed that the ODC gene might act as an oncogene because its overexpression is essential for cell transformation; its activity has been used as a biological marker for evaluating tumor growth and aggressiveness (Deng et al. 2008). Pitkänen et al. (2001) in mice and Lopez-Contreras et al. (2006) in humans showed activity of the mRNA of its paralogue ODC-like gene only in the brain and the testes with no intrinsic decarboxylase activity, concluding that ODC-like acts as a novel antizyme inhibitory protein (AZI2) which can activate ODC by competing for AZ, a family of inhibitory proteins that regulate ODC. However, Jhingran et al. (2012) found the ODC activity associated to the ODC-like protein in the amoeba *Entamoeba histolytica* (Schaudinn) concluding that this protein might have other functions, so far unidentified, including a regulatory role.

ii) Phosphoenolpyruvate cytosolic. The molecular function of phosphoenolpyruvate carboxykinase (PEPCK) (GTP) activity was designated for the sequence described as Phosphoenolpyruvate cytosolic with Blast2GO. The PEPCK is a nucleotide triphosphate (NTP)-dependent enzyme that catalyzes the reversible decarboxylation and the concomitant phosphorylation of oxaloacetic acid to phosphoenolpyruvate (PEP), once considered being the rate-controlling enzyme of gluconeogenesis. Nevertheless, Burgess et al. (2007) suggested that a number of factors in addition to PEPCK activity are responsible for fluctuations in gluconeogenesis. Anyway, an increase in the PEPCK involves an increase in the gluconeogenesis process. Fernandes-Hori et al. (2006) detected an activation of gluconeogenesis in the fish *Brycon cephalus* (Günther) exposed to phenol due to an increasing metabolic energy demand. Also an up-regulation in the

gluconeogenesis was observed in the livers of the zebrafish *Danio rerio* (Hamilton-Buchanan) treated with Mercury (Ung et al. 2010).

iii) Cysteine serine-rich nuclear protein 1-like. A down-regulation in lung, kidney, liver and colon cancers, has been observed for its human orthologue gene (AXUD1) compared with the corresponding normal tissues. This evidence suggested that AXUD1 may have a tumor-suppressor function in those organs (Ishiguro et al. 2001). However, mice generated with deficient CSRNP-1 failed to show any increase in cancer (Gingras et al. 2007). Afterwards, it is important to perform more and specific studies to determine the real function and consequences of the overexpression of both genes (ODC-like and CSRNP-1-like) in *B. microlepidotus.*

iv) Transcription factor Junb-like. Junb is a protein of the JUN protein family which are, together with the FOS protein family, members of the dimeric transcription factor Activator Protein 1 (AP-1). AP-1 has been implicated in many different biological processes, including cell differentiation, proliferation, apoptosis (Passegué et al., 2002) and studies in mouse revealed essential functions of AP-1 in controlling liver development, homeostasis, and disease (Hasenfuss et al. 2014). In the specific case, Junb was pointed out as a negative regulator of cell proliferation, that also upregulates tumor suppressor genes (Shaulian and Karin 2001). Overall, among the upregulated genes detected in the present study, two genes (Transcription factor jun-b-like and CSRNP-1-like) that regulate proteins associated with a tumor suppressor function were found.

On the other hand, two down-regulated genes in the pollution condition were detected, which encode for:

i) Chymotrypsin b-like. This gene was detected as down-regulated in the individuals from the polluted areas; Blast2GO found the molecular function Serine-type peptidase activity which is related with the catalysis of the hydrolysis of peptide bonds in a polypeptide chain.

ii) Probable E3 ubiquitin-protein ligase DTX2-like. Blast2GO identified the molecular function of this gene linked with the zinc ion binding. While it is an interesting gene when studying pollution, in our study, no difference in the zinc concentration between the two conditions was detected (data not shown in this study). Nevertheless, it is important to note that the ubiquitin protein ligases (E3s) are included within a multienzyme system that conjugates ubiquitin to substrate proteins. The first step in the process involves ATP-dependent formation of a thioester between the C terminus of ubiquitin and the active site cystine of ubiquitin activating enzyme (E1). A second thioester intermediate is subsequently created between ubiquitin and one of several ubiquitin-conjugating enzymes (E2s). In the final step of the process, E2s perform together with ubiquitin protein ligases (E3s) to produce an iso-peptide bond between the carboxyl terminus of ubiquitin and a free amino group on the substrate protein. In most cases, ubiquitin conjugation results in multiubiquitin chains that target the substrate for degradation by proteasomes or, in some cases, lysosomes (Kassenbrock and Anderson

2004). Takeyama et al. (2003) reported that the human family of DTX proteins (DTX1, DTX2, and DTX3) function as E3 ligases based on their capacity for self-ubiquitination.

Overall, according to the evidence of the two genes down-regulated in the polluted condition, the organisms inhabiting this condition possibly present lower protein degradation than those of non-polluted places. A similar pattern was observed at the liver of individuals of the fish *Danio rerio* exposed to 21 days of starvation. In these individuals, an overall decrease in metabolic activity, reduced lipid metabolism, protein biosynthesis, proteolysis, and cellular respiration, and increase in gluconeogenesis was observed (Drew et al. 2008). In this study, an increase in gene expression related with the gluconeogenesis and a decrease in expression of genes related with the proteolysis was observed. This raises the possibility that individuals of *B. microlepidotus* inhabiting polluted areas could be exposed to starvation.

In the case of the two SNPs detected, both are located at the mitochondria (genes ND5 and COI), which is the "energy factory" of cells. Therefore, these SNPs were located in a key organelle at the cellular level (Janowsky et al. 2006) that can suffer a number of cytological alterations and apparent changes when exposed to toxic chemicals (Kohler 1999). Evidence of the pollution effects on mitochondria activity has been reported by Padmini et al. (2009), who observed a reduction in the cytochrome c oxidase (CO) activity in fishes exposed to pollution in comparison with individuals not exposed. Our work detected two SNPs with a potential adaptive role to the pollution that are located in the mitochondria, nevertheless this is not the first study with evidence of evolutionary

change at the mitochondria driven by pollution. Belanger-Deschenes et al. (2013) performed a genome scan contrasting individuals of the fish *Perca flavescens* from clean and contaminated lakes. The authors distinguished three SNPs to be putatively under selection: one nuclear (cyclin G1 gene) and two mitochondrial (cytochrome b and NADH dehydrogenase subunit 2 genes). These genes presented allelic correlation with the mean population cadmium concentration. Then, these results pointed out that the two SNPs detected in the mitochondria are potential candidate genes that contribute to the adaptation of the perch in contaminated areas by conferring tolerance of high concentration of metals. In our study, we detected a fixed SNP at the ND5, which is part of the NADH dehydrogenase (complex I). It catalyzes the reduction of ubiquinone with NADH as a reducing substrate which couples with the transfer of protons from the mitochondrial matrix to the intermembranous space. The function of pumping protons across the inner mitochondrial membrane at complex I, II and IV (CO) establishes a proton gradient which can be used for Adenosine Triphosphate (ATP) synthesis (Bao and Li 2011), and mutations in NADH dehydrogenase subunits can affect mitochondrial functions including the ATP production. Furthermore, one FP-DNP SNP was detected in the COI gene, an important component of the complex IV. CO is a key oxidative enzyme in the mitochondria that consumes most of the cellular oxygen and catalyzes the electron transfer to generate ATP via the coupled process of oxidative phosphorylation (Capaldi 1992; Sharpe and Cooper 1998). Hong et al. (2002) reports that the inhibition of CO can cause oxidative stress and mitochondrial-mediated apoptosis. Taking into account all this information, we considered that changes in the ND5 and COI gene could

have an important adaptive effect in the silverside, therefore it is relevant to carry out more detailed studies in this subject.

In conclusion, the pollution in the Maipo River basin is affecting both, the physiological and adaptive levels of the silverside *B. microlepidotus*. These findings provide new insights into our understanding of the species adaptation to the changing environments, including those affected by pollution.

Acknowledgements

JP Oyanedel, Pablo Muñoz Rojas and C Quezada-Romegialli for helping us in the field work and Dr. Arne Nolte for his help in the beginning of this chapter. CVR thanks Conicyt Doctoral Fellowship 21090188, doctoral thesis fellowship 24121005 and doctoral fellowship internship 75130012. Financial support of Fondecyt 1100341, Grant Enlace Universidad de Chile 2013, Basal Grant PFB 023 and ICM P05-002.

REFERENCES

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biology 11: R106.

Bahamondes I, Soto D, Vila I (1979) Hábitos alimentarios de las especies de Atherinidae del Embalse Rapel. Medio Ambiente 4: 3-18.

Bao HG, Li JY (2011) A missense mutation in mitochondrially encoded NADH dehydrogenase subunit-5 gene associates with chicken fetal growth under hypoxia condition. Journal of Animal and Veterinary Advances 10: 1991-1993.

Bélanger-Deschênes S, Couture P, Campbell PGC, Bernatchez L (2013) Evolutionary change driven by metal exposure as revealed by coding SNP genome scan in wild yellow perch (*Perca flavescens*). Ecotoxicology 22: 938-957.

Burgess SC, He TT, Yan Z, Lindner J, Sherry AD, Malloy CR, Browning JD, Magnuson MA (2007) Cytosolic Phosphoenolpyruvate Carboxykinase Does Not Solely Control the Rate of Hepatic Gluconeogenesis in the Intact Mouse Liver. Cell Metabolism 5: 313–320.

Capaldi RA (1992) Structure and function of cytochrome oxidase. Annu Rev Biochem 59: 569–596.

Campos H, Dazarola G, Dyer B, Fuentes L, Gavilan J, Huaquin L, Martinez G, Melendez R, Pequeño G, Ponce F, Ruiz V, Siefeld W, Soto D, Vega R, Vila I (1998) Categorías de conservación de peces nativos de aguas continentales de Chile. Boletín del Museo Nacional de Historia Natural 47: 101-122.

Conesa A, Götz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674–3676.

Cheveruex B, Wetter T, Suhai S (1999) Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB) 99: 45–56.

Comte S, Vila I (1992) Spawning of *Basilichthys microlepidotus* (Jenyns). Journal of Fish Biology 41: 971-981.

Deng W, Jiang X, Mei Y, Sun J, Ma R, Liu X, Sun H, Tian H, and Sun X (2008) Role of ornithine decarboxylase in breast cancer. Acta Biochim Biophys Sin 40: 235-243.

Dirección general de aguas (2004) Diagnostico y clasificación de los cursos y cuerpos de agua según objetivos de calidad, cuenca del Rio Maipo. Ministerio de Obras Públicas, Gobierno de Chile.

Drew ER, Rodnick KJ, Settles M, Wacyk J, Churchill E, Powell MS, Hardy RW, Murdoch GK, Hill RA, Robison BD (2008) Effect of starvation on transcriptomes of brain and liver in adult female zebrafish (*Danio rerio*). Physiol Genomics 35: 283–295

Duarte W, Feito R, Jara C, Moreno C, Orellana AE (1971) Ictiofauna del sistema hidrográfico del río Maipo. Boletín del Museo Nacional de Historia Natural Chile 32: 227-268.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107: 1–15.

Fay JC, Wittkopp PJ (2008) Evaluating the role of natural selection in the evolution of gene regulation. Heredity 100: 191-199.

Fernandes-Hori TS, Marchionni-Avilez I, Kioshi-Inoue L, Moraes G (2006) Metabolical changes induced by chronic phenol exposure in matrinxã *Brycon cephalus* (teleostei: characidae) juveniles. Comparative Biochemistry and Physiology, Part C 143: 67–72

Garcia T, Shen Y, Crawford D, Oleksiak MF, Whitehead A, Walter RB (2012) RNA-Seq reveals complex genetic response to deepwater horizon oil release in *Fundulus grandis*. BMC Genomics 13: 474.

Gingras S, Pelletier S, Boyd K, Ihle JN (2007) Characterization of a Family of Novel Cysteine- Serine-Rich Nuclear Proteins (CSRNP). PLoS ONE 2 (8): e808

Hasenfuss SC, Bakiri L, Thomsen MK, Hamacher R, Wagner EF (2014) Activator Protein 1 Transcription Factor Fos-Related Antigen 1 (Fra-1) Is Dispensable for Murine Liver Fibrosis, but Modulates Xenobiotic Metabolism. Hepatology 59: 261-273.

Hong MY, Chapkin RS, Barhoumi R, Burghardt RC, Turner ND, Henderson CE, Sanders LM, Fan YY, Davidson LA, Murphy ME et al (2002) Fish oil increases mitochondrial phospholipid unsaturation, up-regulating reactive oxygen species and apoptosis in rat colonocytes. Carcinogenesis 23:1919–1925

Huang Q, Dong S, Fang C, Wu X, Ye T, Lin Y (2012) Deep sequencing-based transcriptome profiling analysis of *Oryzias melastigma* exposed to PFOS. Aquatic Toxicology 120-121: 54-58.

Ishiguro H, Tsunoda T, Tanaka T, Fujii Y, Nakamura Y, et al. (2001) Identification of AXUD1, a novel human gene induced by AXIN1 and its reduced expression in human carcinomas of the lung, liver, colon and kidney. Oncogene 20: 5062–5066.

Janowsky BV, Major T, Knapp K, Voos W (2006) The disaggregation activity of the mitochondrial clpB homolog HSP78 maintains HSP70 function during heat stress. J Mol Biol 357:793–807.

Jhingran A , Padmanabhan PK, Singh S, Anamika K, Bakre AA, Bhattacharya S, Bhattacharya A, Srinivasan N, Madhubala R (2008) Characterization of the Entamoeba histolytica Ornithine Decarboxylase-Like Enzyme. PLoS Neglected Tropical Diseases 2: e115.

Kassenbrock CK, Anderson SM (2004) Regulation of Ubiquitin Protein Ligase Activity in c-Cbl by Phosphorylation-induced Conformational Change and Constitutive Activation by Tyrosine to Glutamate Point Mutations. The Journal of Biological Chemistry 279: 28017-28017.

Kohler A (1999) Identification of contaminant induced cellular and subcellular lesions in the liver of flounder, caught at different polluted estuaries. Aquat Toxicol 16: 271–94.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25: 2078–2079.

Lohse M, Bolger A, Nagel A et al. (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Research 40: 622–627.

Lopez-Contreras AJ, Lopez-Garcia C, Jimenez-Cervantes C, Cremades A, Peñafiel R (2006). Mouse Ornithine Decarboxylase-like Gene Encodes an Antizyme Inhibitor

Devoid of Ornithine and Arginine Decarboxylating Activity. The journal of biological chemistry 281: 30896–30906.

Muñoz BM (2007) Cambios en la riqueza íctica del rio Maipo. Tesis Biología con Mención en Medio Ambiente, Facultad de Ciencias, Universidad de Chile.

Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural population. Nature Genetics 32: 261-266.

Olsvik PA, Berg V, Lyche JL (2013) Transcriptional profiling in burbot (*Lota lota*) from Lake Mjøsa-A Norwegian Lake contaminated by several organic pollutants. Ecotoxicology and Environmental Safety 92: 94-103.

Pardo R, Vila I, Bustamante R (2008) Demographic response of *Stratiodrilus aeglaphilus* (Anelida: Histriobdellidae) to organic enrichment: experimental assessment. Hydrobiologia 598: 11-16.

Padmini E, Vijaya Geetha B, Usha Rani M (2009) Pollution induced nitrative stress and heat shock protein 70 overexpression in fish liver mitochondria. Science of the total environment 407: 1307-1317.

Passegué E, Jochum W, Behrens A, Ricci R, Wagner EF (2002) JunB can substitute for Jun in mouse development and cell proliferation. Nature genetics 30: 158-166.

Pitkänen LT, Heiskala M, Andersson LC (2001) Expression of a Novel Human Ornithine Decarboxylaselike Protein in the Central Nervous System and Testes. Biochemical and Biophysical Research Communications 287: 1051–1057

Qiao X, Shi G, Jia R, Chen L, Tian X, Xu J (2012) Physiological and biochemical responses induced by lead stress in *Spirodela polyrhiza*. Plant Growth Regul 67:217–225.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863-864

Sharpe MA, Cooper CE (1998) Interactions of peroxynitrite with mitochondrial cytochrome oxidase: catalytic production of nitric oxide and irreversible inhibition of enzyme activity. J Biol Chem 273: 30961–30972

Shaulian E, Karin M (2001) AP-1 in cell proliferation and survival. Oncogene 20: 2390-2400.

Sotomayor V, Lascano C, Pechen De D'angelo AM, Venturino A (2012) Developmental and polyamine metabolism alterations in *Rhinella arenarum* embryos

exposed to the organophosphate chlorpyrifos. Environmental Toxicology and Chemistry 31: 2052–2058.

Takeyama K, Aguilar RCT, Gu L, He C, Freeman GJ, Kutok JL, Aster JC, Shipp MA (2003) The BAL-binding Protein BBAP and Related Deltex Family Members Exhibit Ubiquitin-Protein Isopeptide Ligase Activity. The journal of Biological Chemistry 278: 21930-21937.

Ung CY, Lam SH, Hlaing MM, Winata CL, Korzh S, Mathavan S, Gong Z (2010) Mercury-induced hepatotoxicity in zebrafish: in vivo mechanistic insights from transcriptome analysis, phenotype anchoring and targeted gene expression validation. BMCGenomics11: 212.

Ungerer MC, Johnson LC, Herman MA (2008) Ecological genomics: understanding gene and genome function in the natural environment. Heredity 100:178–183.

Vega-Retter C, Muñoz-Rojas P, Vila I, Copaja S, Véliz D (2014) Genetic effects of living in a highly polluted environment: the case of the silverside *Basilichthys microlepidotus* (Jenyns) (Teleostei: Atherinopsidae) in the Maipo River basin, central Chile. Population Ecology: DOI 10.1007/s10144-014-0444-3.

Vega-Retter C, Vila I, Véliz D Signatures of directional and balancing selection in the silverside *Basilichthys microlepidotus* (Teleostei: Atherinopsidae) inhabiting a polluted river. Submitted to Journal of Evolutionary Biology.

Véliz D, Catalán L, Pardo R, Acuña P, Diaz A, Poulin E, Vila I (2012) The genus *Basilichthys* (Teleostei : Atherinopsidae) revisited along its Chilean distribution range (21° to 40° S) using variation in morphology and mtDNA. Revista Chilena de Historia Natural 85: 49-59.

World Water Assessment Programme (2009) The United Nations World Water Development Report 3: Water in a Changing World. UNESCO, Paris and Earthscan, London.

Xu Y, Shi GX, Ding CX, Xu XY (2011) Polyamine metabolism and physiological responses of *Potamogeton crispus* leaves under lead stress. Russian Journal of Plant Physiology 58: 460–466.

Yang HY, Shi GX, Li WL, Wu WL (2013) Exogenous spermidine enhances *Hydrocharis dubia* Cadmium tolerance. Russian Journal of Plant Physiology 60: 770–775.

GENERAL DISCUSSION

Pollution is currently a main factor affecting natural populations in freshwater systems (e.g., Dudgeon et al. 2006; Strayer and Dudgeon 2010), being especially important in rivers surrounded by large cities around the world. In our country, rivers have been impacted by the use of water for sewage and mining exploitation, thus, pollution is present in most rivers. In central Chile, where half of the Chilean population inhabits, the Maipo River basin showed to be affected by the human activity presenting areas with high pollution. Considering this information, the main goal of this research was to determine the genetic consequences of the pollution on populations of the silverside *Basilichthys microlepidotus* (Jenyns) in this basin. This study focus in three specific objectives i) to determine the relationship of pollution with the genetic structure of *B. microlepidotus* by comparing the gene diversity and migration rates of its populations and, ii) to determine the relationship between pollution and allele selection, this by performing a genome scan and iii) to determine the relationship of the gene expression and genetic variability with the pollution.

In the first chapter, the population structure of *B. microlepidotus* was studied. In total, five populations of *B. microlepidotus* were detected, two populations inhabiting polluted areas and three inhabiting non-polluted areas. Further, this study did not find evidence that pollution affects silverside populations, indicating that this species is able to tolerate the pollution of the Maipo River basin. In this context, literature evidenced that some

fish species are specially affected by water pollution, other studies showed species that are tolerant to a contaminated areas. Therefore, the present study supports the idea that different species react in a different way to face the pollution. This research also contributes to our understanding on native species conservation by pointing out that not all native species answer in the same way to pollution making necessary case by case studies to determine an adequate conservation plan.

In the second chapter, the proportion of loci under directional and balancing selection related with the pollution was studied. This research points out the presence of signatures of directional selection related to the pollution in *B. microlepidotus* inhabiting the Maipo River basin. A total of six loci (4% of the total loci) were detected as under directional selection due to pollution, three loci (2% of the total loci) for each contaminated population. As an example, a similar pattern was detected by Williams and Oleksiak (2008), which found that 1% to 6 % of the AFLP loci of *Fundulus heteroclitus* populations exposed to chronic contamination could be under selection or linked to areas of the genome that are under directional selection due to pollution. Thus, pollution could be linked with changes in the allelic frequencies when populations of *B. microlepidotus* inhabiting polluted and non-polluted areas are compared. Further, balancing selection was detected in four loci (2.7% of the total). To our knowledge, this is the first study that identifies loci under balancing selection related to pollution, and therefore it indicates that pollution could be responsible for maintaining polymorphism in the polluted populations.

Finally, in the last chapter of this research, genes with differential expression and the presence of SNPs between the polluted and non-polluted condition were studied. After the analysis that imposed several restrictions to avoid false positive results, a total of six genes were found with differential expression, two down-regulated and four up-regulated in individuals sampled in the polluted areas. Also, two candidates SNPs were detected. Among the up-regulated genes in polluted areas, one gene was related with the regulation of the gluconeogenesis process and the two genes down-regulated were related with the protein degradation. This evidence suggests that pollution affects food availability for *B. microlepidotus*; the same pattern was observed in the fish *Danio rerio* exposed to 21 days of starvation (Drew et al. 2008). Moreover, we found that among the genes up-regulated in the polluted condition, two genes were possibly related with tumor suppression and one gene showed to be associated with cancer development; this finding makes necessary to perform more studies to clarify this discovery. The two candidate SNPs were both located at mitochondrial genes. Considering the mitochondria as the "energy factory" of cells and a key at the cellular level (Janowsky et al. 2006) these two SNPs could have an important adaptive effect in the silverside; it is relevant to carry out more detailed studies in this subject.

At our knowledge, this study is the first and most complete analysis realized on the effects of the pollution in a native freshwater fish in Chile using a genetic approach. Here I demonstrate the genetic, selective and physiological effects of pollution, being this study the initial step towards the understanding of the multiple genetic effects related with the human activity in the freshwater environments. However, it is necessary

to take into account the limitations of this study. First, the no effects of the pollution on the gene diversity and migration rate could be related with the number of markers analyzed. Although eight microsatellites were analyzed, it is necessary to test with more markers in order to confirm our discovery. Further, each site inside the categories polluted and non-polluted used in this analysis have particular physicochemical characteristics, which could have influenced the results. Although I think that is an important point to inform, it is impossible to separate the effects of each pollutant in a field experiment.

In conclusion, this study found that *B. microlepidotus* was tolerant to the pollution present in the Maipo River basin. Associated to this finding, the pollution acts as a selective driver by producing changes in allele frequencies different to this expected under gene drift. Interestingly, evidence of balancing selection related with pollution was first found in this study. Further, pollution also affects *B. microlepidotus* in a physiological level by changing its genic expression profile and probably in a genetic level by selecting different alleles, when individuals inhabiting polluted and non-polluted sites in the Maipo River basin are compared. Overall, pollution is producing a selective, physiological and adaptive effect in the populations of *B. microlepidotus* inhabiting the Maipo River basin.

REFERENCES

Drew ER, Rodnick KJ, Settles M, Wacyk J, Churchill E, Powell MS, Hardy RW, Murdoch GK, Hill RA, Robison BD (2008) Effect of starvation on transcriptomes of brain and liver in adult female zebrafish (*Danio rerio*). Physiol Genomics 35: 283–295

Dudgeon D, Arthington AH, Gessner MO, Kawabata Z-I, Knowler DJ, Lequeve C, Naiman RJ, Prieur-Richard A-H, Soto D, Stiassny MLJ, Sullivan CA (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. Biol Rev 81:163–182.

Janowsky BV, Major T, Knapp K, Voos W (2006) The disaggregation activity of the mitochondrial clpB homolog HSP78 maintains HSP70 function during heat stress. J Mol Biol 357:793–807.

Strayer DL, Dudgeon D (2010) Freshwater biodiversity conservation: recent progress and future challenges. J N Am Benthol Soc 29:344–358

Williams LM, Oleksiak MF (2008) Signatures of selection in natural populations adapted to chronic pollution. BMC Evol Biol 8:282–293

Annexes

Annex 1. Summary of the physicochemical parameters measured at each sampling site. Each measure is the average of three samples. EC: electrical conductivity, TDS: total dissolved solids, DO: dissolved oxygen, SO2-4: sulfate, NO-2: nitrite, Cl-: chloride, NH+4: ammonium, NO-3: nitrate, PO3-4: phosphate, Na+: sodium, K+: potassium, Ca2+: calcium, Mg2+: magnesium, Cu: copper, Mo: molybdenum, BOD5: Biological Oxygen Demand. ***: under limit detection.

Site	Date	EC (µS/cm)	pH	TDS (ppm)	DO (µg/L)	NO^-_2 (µg/L)	SO^{2-}_4(mg/L)	NH^+_4 (µg/L)	NO^-_3 (µg/L)	PO^{3-}_4 (µg/L)	Na^+ (mg/L)	K^+ (mg/L)	Ca^{2+} (mg/L)	Mg^{2+} (mg/L)	BOD_5 (mg/L)	Cl^- (mg/L)	Mo (µg/L)	Cu (µg/L)
SFM	01/2011	402.00	7.18	220.00	6.53													
	09/2011	440.33	7.18	221.00	12.24	0.51	27.38											
	07/2012	405.00	7.04	203.00	11.01	0.69	27.77	0.00	551.56	12.56	13.40	0.88	37.59	12.75	2.58	33.33	***	3.62
	11/2012	403.50	6.62	199.50	9.04	0.09	22.80	32.00	594.17	20.56	19.46	2.10	56.27	11.48	2.40	30.83		
PU	01/2011	316.00	7.16	152.00	5.39													
	05/2011	349.67	7.29	175.00														
	09/2011	330.00	7.58	130.00	11.01	0.68	7.69											
	12/2012	300.00	7.30	149.00	8.97	0.00	7.11	323.88										
	07/2012	203.00	7.32	103.33	12.24	2.17	5.48	0.00	109.93	4.19	9.21	0.00	26.87	8.42	2.72	19.33	1.15	0.68
PN	01/2011	1273.00	7.42	636.00	6.94													
	09/2011	1171.67	7.15	581.33	7.41	0.76	97.85											
	07/2012	1178.00	6.82	588.67	7.89	5.20	92.49	0.44	309.34	5.18	58.03	1.90	130.37	29.33	2.00	160.00	0.75	1.14
PEL	01/2011	1248.00	7.20	621.00	11.02													
	09/2011	1447.33	7.93	723.67														
	01/2012	1273.00	7.27	636.00	8.77	109.99	94.14	242.24										
	07/2012	1278.00	7.24	638.33	10.33	49.08	98.14	20.87	771.39	15.96	15.23	2.36	132.59	30.65	3.30	156.00	0.32	2.28
	11/2012	1282.50	7.33	638.50	9.76	52.74	67.05	54.13	814.08	25.82	76.26	5.71	207.40	36.38	4.09	160.83		
MEL	01/2011	1360.00	7.35	680.00	7.75													
	04/2011	1381.67	7.23	691.00														
	05/2012	1424.00	7.09	711.33	5.92	0.00	114.03	5.17	1053.88	17.77	104.90	5.65	144.42	31.49	1.49	187.99	***	0.01

	11/2012	1101.00	7.36	558.67	7.29	200.87	62.27	611.45	445.83	93.30	84.23	5.28	174.35	22.43	3.90	150.83		
IM	01/2011	1129.00	7.19	565.00	11.22													
	09/2011	1129.67	7.23	563.00	7.54	9.43	91.90											
	05/2012	1139.00	7.03	569.00	7.48	1.75	91.22	1.16	1787.01	3.21	83.65	4.34	121.09	17.43	1.84	137.33	2.19	2.09
	11/2012	1256.00	7.19	628.50	7.03	0.06	70.07	12.55	930.75	4.10	76.97	3.86	201.53	22.30	0.89	152.50		
MA	01/2012	1229.00	7.58	614.00	7.89	8.81	101.83	47.27										
	05/2012	1259.00	7.71	628.33	10.27	11.75	99.11	5.05	904.79	24.32	115.39	6.00	121.85	21.87	3.94	170.00	2.30	3.82

Annex 2. Summary of the genetic variables estimated from the eight microsatellites analyzed for *B. microlepidotus* at seven sites. *n*: sampling size, N_A: number of alleles, H_E: expected heterozygosity, H_O: observed heterozygosity, F_{IS}: inbreeding coefficient and the respective *P* value estimated with permutation analysis, bold values of *P* indicate deviation from the HWE. *: Deviation from the HWE in the total F_{IS}. r_{xy}: relatedness degree obtained with the IDENTIX software, with statistical significance estimated using permutation method.

	SFM	MA	PU	PN	PEL	MEL	IM
Odon01							
n	24	24	24	19	23	22	20
N_A	3	4	3	4	4	4	3
H_E	0.442	0.568	0.595	0.679	0.698	0.569	0.549
H_O	0.375	0.619	0.667	0.643	0.727	0.526	0.444
F_{IS}	0.172	-0.066	-0.096	0.089	-0.018	0.102	0.247
P	0.281	0.447	0.380	0.384	0.560	0.375	0.239
Odon02							
n	24	24	24	19	23	22	20
N_A	2	3	3	3	2	1	2
H_E	0.078	0.081	0.223	0.204	0.043	0.000	0.266
H_O	0.082	0.082	0.250	0.111	0.044	0.000	0.105
F_{IS}	-0.022	-0.011	-0.099	0.477	-0.001	-	0.621
P	1.000	1.000	0.704	0.068	1.000	-	**0.032**
Odon07							
n	24	24	24	19	23	22	20
N_A	4	3	4	4	3	3	3
H_E	0.659	0.574	0.572	0.5414	0.588	0.575	0.590
H_O	0.421	0.571	0.632	0.462	0.522	0.667	0.412
F_{IS}	0.384	0.041	-0.077	0.186	0.136	-0.136	0.329
P	**0.015**	0.535	0.445	0.277	0.271	0.848	0.069
Odon09							
n	24	24	24	19	23	22	20
N_A	6	7	6	5	4	6	7

H_E	0.765	0.654	0.694	0.579	0.538	0.675	0.741
H_O	0.583	0.696	0.708	0.231	0.391	0.727	0.611
F_{IS}	0.257	-0.041	0.000	0.626	0.293	-0.062	0.203
P	**0.013**	0.485	0.594	**0.001**	0.053	0.425	0.091
Odon19							
n	24	24	24	19	23	22	20
N_A	6	6	5	4	1	5	5
H_E	0.297	0.332	0.296	0.281	0.000	0.171	0.560
H_O	0.333	0.333	0.250	0.263	0.000	0.181	0.421
F_{IS}	-0.102	0.016	0.176	0.091	-	-0.037	0.272
P	0.500	0.600	0.193	0.371	-	1.000	0.056
Odon39							
n	24	24	24	19	23	22	20
N_A	2	2	2	2	2	2	2
H_E	0.239	0.153	0.349	0.45	0.194	0.245	0.415
H_O	0.167	0.056	0.150	0.474	0.217	0.191	0.353
F_{IS}	0.328	0.653	0.587	-0.025	-0.100	0.245	0.179
P	0.274	0.086	**0.022**	0.665	0.783	0.339	0.436
Odon59							
n	24	24	24	19	23	22	20
N_A	2	1	3	3	2	4	4
H_E	0.187	0.000	0.155	0.656	0.227	0.543	0.681
H_O	0.208	0.000	0.167	0.500	0.261	0.600	0.444
F_{IS}	-0.095	-	-0.051	0.264	-0.128	-0.080	0.371
P	0.790	-	0.870	0.092	0.690	0.436	**0.018**
Odon71							
n	24	24	24	19	23	22	20
N_A	4	3	2	2	3	2	3
H_E	0.595	0.586	0.278	0.145	0.232	0.420	0.493
H_O	0.391	0.435	0.333	0.158	0.261	0.400	0.412
F_{IS}	0.361	0.279	-0.179	-0.058	-0.104	0.073	0.194
P	**0.021**	0.056	0.500	1.000	0.689	0.569	0.192
Total F_{IS}	0.237*	0.078	0.025	0.229*	0.061	-0.006	0.284*
r_{xy}	-0.035	-0.027	-0.054	-0.054	-0.052	-0.095	-0.045
P	0.150	0.060	0.601	0.223	0.560	0.990	0.060

Annex 3. Current number of effective migrants per generation (N_em) estimated between pairs of populations of *B. microlepidotus* within the Maipo River basin.

From/To	MEL	IM-PN	MA-SFM	PU	PEL
MEL	21.665	1.593	0.224	0.255	0.507
IM-PN	0.819	105.95	0.256	0.659	0.634
MA-SFM	1.039	2.048	27.426	1.893	0.761
PU	0.409	1.366	1.121	15.867	1.141
PEL	7.526	2.731	3.012	2.595	39.208

Annex 4. Historical number of effective migrants per generation (N_em) estimated between pairs of populations of *B. microlepidotus* within the Maipo River basin.

From/To	MEL	IM-PN	MA-SFM	PU	PEL
MEL		2.395	4.135	1.686	0.861
IM-PN	0.468		0.411	2.367	0.381
MA-SFM	1.043	1.996		1.461	0.309
PU	0.444	2.761	1.279		0.077
PEL	0.939	3.238	2.642	2.014	