Polar Biol
DOI 10.1007/s00300-013-1370-0

ORIGINAL PAPER

Putative selected markers in the *Chionodraco* genus detected by interspecific outlier tests

Cecilia Agostini · Chiara Papetti · Tomaso Patarnello · Felix C. Mark · Lorenzo Zane · Ilaria A. M. Marino

Received: 14 March 2013 / Revised: 20 June 2013 / Accepted: 27 June 2013

Abstract The identification of loci under selection (outliers) is a major challenge in evolutionary biology, being critical to comprehend evolutionary processes leading to population differentiation and speciation, and for conservation purposes, also in light of recent climate change. However, detection of selected loci can be difficult when populations are weakly differentiated. This is the case of marine fish populations, often characterized by high levels of gene flow and connectivity, and particularly of fish living in the Antarctic marine environment, characterized by a complex and strong circulating system promoting individual dispersal all around the continent. With the final aim of identifying outlier loci putatively under selection in the *Chionodraco* genus, we used 21 microsatellites, including both genomic (Type II) and EST-linked loci (Type I), to investigate the genetic differentiation among the three recently derived *Chionodraco* species that are endemic to the freezing Antarctic waters. Neutrality tests were applied in interspecific comparisons in order to identify candidate loci showing high levels of genetic differentiation, which might reveal imprints of past selection. Three outlier loci were identified, detecting a higher differentiation between species than did neutral loci. Outliers showed sequence similarity to a calmodulin gene, to an antifreeze glycoprotein/trypsinogen-like protease gene and to nonannotated fish mRNAs. Selective pressures acting on outlier loci identified in this study might reflect past evolutionary processes, which led to species divergence and local adaptation in the *Chionodraco* genus. Used loci will provide a valuable tool for future population genetic studies in Antarctic notothenioids.

Keywords Genome-wide selection scan · Genetic differentiation · Local adaptation · EST-linked microsatellites · Standardized F_{ST} · Antarctic icefish

Electronic supplementary material The online version of this article (doi:10.1007/s00300-013-1370-0) contains supplementary material, which is available to authorized users.

C. Agostini · C. Papetti · L. Zane (✉) · I. A. M. Marino
Department of Biology, University of Padova, Padua, Italy
e-mail: lorenzo.zane@unipd.it

T. Patarnello
Department of Comparative Biomedicine and Food Science, University of Padova, Agripolis, Legnaro, Padua, Italy

F. C. Mark
Integrative Ecophysiology, Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany

Introduction

The idea that marine species are genetically homogeneous throughout their range of distribution is traditionally assumed. Large population sizes and wide spatial distributions, associated with an extensive potential for dispersal, in the absence of evident barriers to gene flow, are supposed to limit genetic structuring among local populations (Ward et al. 1994). Typically, genetic studies of natural populations employ neutral molecular markers, like SNPs and microsatellites, which permit to elucidate various aspects of species biology. These loci also allow the estimation of demographic parameters, such as effective population size and migration rate, the inference of which would be biased by the effects of natural selection acting in a locus-specific manner (Avise 1994). The growing interest in understanding the genetic bases of ecologically important traits and in studying local adaptation has, however, gradually shifted the attention to genetic markers

Published online: 14 July 2013

influenced by natural selection, as revealed by the development of several methods aimed at detecting putative selected loci from genome-wide scans (Bowcock et al. 1991; Beaumont and Nichols 1996; Vitalis et al. 2001; Schlötterer 2002; Porter 2003; Beaumont and Balding 2004; Foll and Gaggiotti 2008).

In fish, genome scan studies accomplished so far, in which many individuals were screened using randomly selected molecular markers, reported very few instances of outlier detection; these cases include several members of the Salmonidae family (e.g., Freamo et al. 2011; Seeb et al. 2011), which are well known to be highly structured, some coral reef fishes (e.g., Fauvelot et al. 2007), and a number of other economically important species like Atlantic cod (e.g., Nielsen et al. 2009) and Atlantic herring (e.g. André et al. 2010). Difficulties in finding selected loci in marine fishes may be referred to the limited power of available programs to identify outlier loci when population differentiation is weak. Most of the existing statistical methods for detecting recent signatures of selection from allele frequency data are based on the original idea developed by Lewontin and Krakauer (1973), according to which loci showing a significantly higher or lower level of genetic differentiation than expected under an appropriate neutral population-genetics model can be considered candidates for being influenced by directional (adaptive) or balancing selection, respectively. The level of genetic differentiation is typically described by the F_{ST} index, which summarizes differences in allele frequency distribution between populations. The main difficulty of these methods is to obtain by simulations the expected F_{ST} distribution under neutrality, which should depict the expected variance of F_{ST} values across loci with different levels of genetic variability. From a statistical standpoint, outlier detection is particularly difficult when population structuring is weak and then neutral mean F_{ST} values are low (Beaumont and Balding 2004; Foll and Gaggiotti 2008).

The identification of outlier loci in marine fish is particularly challenging when environmental settings favor gene flow across populations with a consequent homogenizing effect. This is the case of Antarctic waters, where the complex ocean circulating system, which promotes dispersal and connectivity all around Antarctica, contributes to limit genetic differentiation among populations (Matschiner et al. 2009; Papetti et al. 2009, 2012 Damerau et al. 2012).

In this study, we analyzed the genetic differentiation between the three recently derived species of the *Chionodraco* genus (Notothenioidei, Channichthyidae), namely *Chionodraco hamatus*, *Chionodraco myersi,* and *Chionodraco rastrospinosus*, using a panel of 21 microsatellite loci. The three species, which are endemic to the Southern Ocean and whose taxonomical status is based on few morphological differences and molecular data (Fischer and Hureau 1985; Patarnello et al. 2003), are thought to have diverged between 2 and 1.8 millions of years ago (additional result not reported in Near et al. 2012). Furthermore, we performed an F_{ST}-based survey to search for loci showing a high level of genetic differentiation between the three *Chionodraco* species and thus possibly influenced by selective forces acting during their shallow evolutionary history. To allow future investigations at the population level, we used a panel of loci, including both microsatellites originally isolated from genomic DNA and EST-linked loci, that cross-amplified in all the three species. Cross-amplified molecular markers, such as microsatellites, have proven to be valuable tools for species discrimination and individual assignment to nominal species also in the notothenioid family Trematominae (Van de Putte et al. 2009).

Materials and methods

Sample collection and DNA extraction

Population samples of *C. hamatus*, *C. myersi,* and *C. rastrospinosus* were collected between 1988 and 2007 at four different locations: the Weddell Sea, the Ross Sea (Terra-nova Bay), and the Antarctic Peninsula (Elephant Island and Joinville Island) (Table 1). A small piece of muscle tissue was collected from each specimen and preserved at −80 °C or in ethanol 90 % until molecular analysis.

Total genomic DNA for each individual was extracted from 10 to 100 mg of muscle tissue following a standard salting out protocol (Patwary et al. 1994). DNA solutions were stored at −20 °C before PCR amplification.

Genetic and statistical analysis: DNA amplification and genotyping

Twenty-one microsatellite loci were amplified in 108 specimens: 10 loci originally isolated from *C. rastrospinosus*, *Chaenocephalus aceratus,* and *Pleuragramma antarcticum* genomic DNA (Papetti et al. 2006, 2011; Susana et al. 2007), and 11 EST-linked loci isolated from about 24,000 contigs obtained by a high-throughput sequencing of a normalized cDNA library from *C. hamatus* muscle (Molecular Ecology Resources Primer Development Consortium et al. 2011; Coppe et al. 2013). According to O'Brien et al. (1993), microsatellite loci isolated from genomic DNA, which are anonymous genomic sequences, will be called Type II loci, while EST-linked loci, which are found inside or flanking coding gene sequence, will be defined as Type I loci. For the sake of simplicity, final conditions for all the loci used in this study are reported in Table S1 (supplementary materials).

Table 1 Summary of samples analyzed in this study

Species	Population sample	Collection cruise	Sample acronym	Sample size
C. hamatus	Weddell Sea	ANT-VII/4[a]	ChWS88	9
	Ross Sea (Terranova Bay)	11th Italian expedition PNRA[b]	ChRS95	23
C. myersi	Ross Sea (Terranova Bay)	5th Italian expedition PNRA[c]	CmRS89	27
	Weddell Sea	ANT-XXI/2[d]	CmWS03	10
C. rastrospinosus	Elephant Island	ANT-XIV/2[e]	CrEI96	19
	Joinville Island	ANT-XXIII/8[f]	CrJI06	20

Reported are as follows: species name, site of collection of each population sample, collection cruise when sampling was performed, sample acronym used in this paper, and sample size

[a] ANTARKTIS expedition, RV "Polarstern" ANT-VII/4 (EPOS leg3), Weddell Sea, 1988/1989, AWI

[b] 11th Italian expedition, Ross Sea, Terranova Bay, 1995/1996, PNRA

[c] 5th Italian expedition, Ross Sea, Terranova Bay, 1989/1990, PNRA

[d] ANTARKTIS expedition, RV "Polarstern" ANT-XXI/2 (Bendex), Weddell Sea, 2003/2004, AWI

[e] ANTARKTIS expedition, RV "Polarstern" ANT-XIV/2, Antarctic Peninsula, 1996/1997, AWI

[f] ANTARKTIS expedition, RV "Polarstern" ANT-XXIII/8, Joinville Island, 2006/2007, AWI

Fragment analysis was performed on an ABI 3730xl automated sequencer, and microsatellite analysis was carried out using PEAK SCANNER version 1.0 (Applied Biosystems). In order to minimize the negative consequences of a poor allele calling, binning was automated with the software FLEXIBIN version 2 (Amos et al. 2007) and the final scoring was then manually checked to ensure the accuracy of the process. MICRO-CHECKER version 2.2.3 (Van Oosterhout et al. 2004) was used to test for null alleles, stuttering and large allele dropout presence, and results for loci showing null alleles were corrected using FREENA (Chapuis and Estoup 2007).

Linkage disequilibrium, Hardy–Weinberg equilibrium (HWE), and genetic diversity

Descriptive analyses, such as number of alleles (N_A), observed heterozygosity (H_{Obs}), and unbiased expected heterozygosity (H_{Exp}), were computed for each locus using GENETIX version 4.05.2 (Belkhir et al. 1996–2004), while allelic richness (A_R), the number of alleles independent of the sample size, was calculated with FSTAT version 2.9.3 (Goudet 2001). The software GENEPOP version online 4.0.10 (Raymond and Rousset 1995; Rousset 2008) was used to test for HWE and for genotypic linkage equilibrium between pairs of loci in each species. Significance of all tests was estimated by the Markov Chain method (dememorization number = 10,000; number of batches = 500; number of iterations per batch = 10,000). When needed, significance threshold ($\alpha = 0.05$) was adjusted using a standard Bonferroni correction for multiple tests (Rice 1989).

Genetic differentiation

Genetic differentiation between species and population samples was quantified by computing pairwise estimators of F_{ST} with FSTAT version 2.9.3 according to Weir and Cockerham (1984). The 95 % CI were estimated by 15,000 bootstrap replicates over loci, and p values were assessed by 1,000 permutation tests in GENETIX version 4.05.2 (Belkhir et al. 1996–2004). Statistical significance level was adjusted, when needed, against type I errors using a standard Bonferroni correction (Rice 1989). The statistic was performed for the whole set of 21 loci, for the same set after the exclusion of outlier loci found by neutrality tests (see below), and separately for Type I and Type II loci, excluding outliers, in order to notice potential differences in the level of detected genetic differentiation. Type I loci, situated in transcribed regions of the genome, are expected to show low levels of variability and to be highly conserved for the presence of functional constraints and of the effects of purifying selection. In contrast, Type II loci are expected to be highly polymorphic because microsatellites are more frequent in noncoding regions of the genome and, as a consequence, they have a higher probability to be neutral markers. Standardized measures of pairwise F_{ST} (F'_{ST}) (independent of the heterozygosity level) were calculated to facilitate the comparison of results obtained by the two sets of loci, which are expected to differ in heterozygosity values and hence in maximum levels of detectable genetic divergence (Hedrick 2005). The calculation was performed by dividing the original F_{ST} estimate by its maximum value, which was obtained using the recoded data file with the program RECODEDATA version 0.1 (Meirmans 2006).

Neutrality tests

To search for the signature of selection, we applied two different neutrality tests relying on the expectation that spatially varying divergent selection should increase the observed level of genetic differentiation between populations/species. Nevertheless, since the two tests are based on different assumptions and parameters, the concordant detection of outlier loci with more than one statistical approach will reduce the number of false positives (Vasemagi et al. 2005). Both tests were applied in all pairwise comparisons between the three species, while they were not used in intraspecific comparisons because of the limited sample size of the individual population samples.

The first method, implemented in the software LOSITAN (Antao et al. 2008), uses coalescent simulations based on the observed data and an island model of migration to generate the expected distribution of F_{ST} versus H_{Exp} with neutral markers. This distribution is then used to identify outlier loci detecting significantly high or low F_{ST} values compared to neutral expectations. These outlier loci are candidates for being subject to natural selection (Antao et al. 2008). In our analysis, 1,000,000 simulations were run assuming a stepwise mutation model (SMM). To better estimate the status of each locus, all pairwise comparisons between species were performed considering firstly Type II loci and then the same set of loci (excluding those eventually resulting under selection) plus one Type I locus at a time. This approach was adopted because the two sets of loci showed different ranges of F_{ST} and H_{Exp} values and because the program failed to simulate the F_{ST} versus H_{Exp} expected distribution when analyzing the two sets together. Moreover, given the higher probability of EST-linked loci to be under selection, considering these loci one by one would not affect the simulation process of the F_{ST} versus H_{Exp} expected distribution in a neutral scenario. In contrast, since microsatellites are more frequent in noncoding regions of the genome, we expected loci isolated from genomic DNA to be neutral markers and, as a consequence, not to affect the simulation process.

The second approach for detecting selection, implemented in the software package DETSEL (Vitalis 2003), simulates a coalescent process of divergence of two populations from a common ancestral population. The program identifies putative outlier loci relying on the population-specific parameters of population divergence, F_1 and F_2, which are simple functions of the parameters of interest of the model. The expected distributions of F_1 and F_2 were generated maintaining the same number of allelic states as in the observed data and using different combinations of the following nuisance parameters: mutation rate under infinite allele model (IAM) 0.005, 0.001, and 0.0001; ancestral population size 500, 1,000, and 10,000; population size before the split 50 and 500; time since an assumed bottleneck event 50, 100, and 200 generations; and time since the population split 50 and 100 generations. For each pairwise comparison, 100,000–500,000 coalescent simulations were performed. All loci lying outside the 99 % probability region, simulated in a neutral scenario, were considered candidates for being subject to natural selection. The same approach used in LOSITAN was chosen to improve the coalescent simulation process (all Type II, excluding those eventually resulting under selection, plus one Type I locus at a time).

Results

Microsatellite genetic variability

All 21 microsatellite loci were successfully amplified in the three *Chionodraco* species. One-way analysis of variance (ANOVA) showed no significant differences in H_{Obs}, H_{Exp}, and A_R across sampling locations in each of the three species (p value >0.05). All loci proved to be polymorphic. For the 10 Type II loci, N_A per locus in the three species ranged from 2 (Ca48 in *C. myersi*) to 32 (Cr171 in *C. rastrospinosus*), while N_A for the 11 Type I loci ranged from 2 (Ch126 in *C. hamatus*, Ch1968 in *C. myersi*, and Ch8501 in *C. hamatus* and *C. myersi*) to 11 (Ch8461 in *C. myersi*) (Tables S2 and S3 supplementary materials). Mean N_A for the two sets of loci in the three species (17.40 combining Type II loci and 7.36 considering Type I loci) was significantly different (p value <0.05 using one-way ANOVA) indicating a lower level of allelic variability at Type I loci. A significant decrease in variability was detected at Type I loci also when A_R, the number of alleles independent of the sample size, was considered (Table 2).

At Type II loci, H_{Obs} in the three species ranged from 0.1892 (Ca21 in *C. myersi*) to 0.9444 (Cr127 in *C. myersi*), while H_{Exp} ranged from 0.2118 (Cr38 in *C. rastrospinosus*) to 0.9710 (Cr171 in *C. rastrospinosus*) (Table S2 supplementary materials). A significant excess of homozygotes

Table 2 Mean values and standard deviations for number of alleles (N_A), allelic richness (A_R), observed (H_{Obs}) and unbiased expected heterozygosity (H_{Exp}) at 10 Type II loci and at 11 Type I loci

	10 Type II loci		11 Type I loci		p value (ANOVA)
	Mean	SD	Mean	SD	
N_A	17.40	11.19	7.36	3.04	0.0100
A_R	12.02	7.33	5.55	2.10	0.0110
H_{Obs}	0.5589	0.0200	0.4397	0.0500	<0.0001
H_{Exp}	0.6842	0.0300	0.4653	0.0500	<0.0001

p values resulted from one-way ANOVA are reported

was found at several loci: Cr15 in *C. hamatus* and *C. rastrospinosus*, Cr171 in *C. rastrospinosus*, Cr236 in *C. hamatus*, Ca21 in *C. hamatus* and *C. myersi*, and Ca86 in *C. myersi*. At Type I loci, H_{Obs} in the three species ranged from 0.0313 (Ch126 and Ch8501 in *C. hamatus*) to 0.8649 (Ch8461 in *C. myersi*), while H_{Exp} ranged from 0.0313 (Ch126 and Ch8501 in *C. hamatus*) to 0.8482 (Ch2309 in *C. hamatus*) (Table S3 supplementary materials). A significant excess of homozygotes was found at locus Ch5817 in *C. myersi*. A significant difference was detected, using one-way ANOVA, when comparing mean H_{Obs} and H_{Exp} for the two sets of loci in the three species (p value <0.05) (Table 2). This result confirmed a differential level of genetic variability at Type I loci, which showed significantly lower values of N_A, A_R, H_{Obs}, and H_{Exp} compared to Type II loci, suggesting the presence of functional constraints and of the effects of purifying selection.

Hardy–Weinberg equilibrium probabilities were calculated for each locus and for each species because of the genetic homogeneity observed in population samples within species. A significant departure from HWE was found at 7 out of 30 tests at Type II loci and at 1 out of 33 tests when considering Type I loci. In all the tests showing a departure from HWE, a significant deficit of heterozygotes was found and MICRO-CHECKER version 2.2.3 indicated the presence of null alleles. Subsequent analyses were performed both with the original and the corrected dataset suggested by MICRO-CHECKER. Furthermore, the excluding null alleles (ENA) correction method implemented in FREENA was carried out to correct for the positive bias induced by the presence of null alleles on F_{ST} estimation. Since both the corrections performed yielded comparable results, the original dataset was maintained for successive analyses. Overall, tests for linkage disequilibrium among all loci showed no significant departures from expected values (p value >0.01).

Genetic differentiation

Pairwise estimators of F_{ST} were calculated both between populations of each species and between species, which are characterized by a limited morphological and genetic differentiation (Fischer and Hureau 1985; Patarnello et al. 2003). The analysis was performed for the complete set of 21 loci, for the 18 neutral loci detected by neutrality tests (see below) and separately for Type I and Type II loci excluding the outlier loci Cr38, Ch684, and Ch8501. For all these sets of loci, pairwise F_{ST} between population samples within species were close to zero and not statistically significant (data not shown). For the limited sample size of individual population samples, and for the genetic homogeneity found in intraspecific comparisons, samples of the same species were grouped. Conversely, pairwise F_{ST} between species were highly significant (p value <0.0001) whatever set of loci was considered (Table 3). When considering all 21 loci, a lower level of differentiation between *C. hamatus* and *C. rastrospinosus* was found; this pattern persisted considering the standardized measure of pairwise F_{ST} (F'_{ST}) and after the removal of the three outlier loci suggested by neutrality tests. In all pairwise comparisons between species, F_{ST} values decreased when outlier loci were excluded from the analysis. Single-locus F_{ST} and F'_{ST} values at outlier loci (Cr38, Ch684, Ch8501) detected a high level of genetic differentiation, specifically in those pairwise comparisons in which they were identified as outlier by neutrality tests (Table 4); in these comparisons, F_{ST} and F'_{ST} estimates clearly exceeded the respective F_{ST} and F'_{ST} 95 % CI between species calculated with 18 putative neutral loci (Table 5).

Neutrality tests

Two different neutrality tests, based on different assumptions and parameters, were applied to detect outlier loci. Simulation results from both tests identified as outlier loci Cr38 and Ch684, in all pairwise comparisons including *C. rastrospinosus*, and locus Ch8501 in all pairwise comparisons including *C. myersi* (p value <0.01). These three loci fell outside the upper 99 % CI of the F_{ST} versus H_{Exp} expected distribution obtained with LOSITAN and lay outside the 99 % probability region of population-specific parameters, F_1 and F_2, simulated by DETSEL; thus, they are potential candidates for being subject to directional selection (two exemplifying results are shown in Figs. 1, 2).

Discussion

The identification of outlier loci in intraspecific comparisons is very challenging in marine fish because of the low levels of population genetic structuring. This task can be even more difficult when oceanographic settings favor individual dispersal and gene flow among populations; this is the case of the Antarctic marine environment, characterized by a complex system of currents. For these reasons and with the aim of detecting outlier loci putatively under selection in the *Chionodraco* species, we moved from the population level to the species level and investigated the pattern of genetic differentiation among the three *Chionodraco* species by using 21 microsatellite loci. Specifically, we used both EST-linked loci (Type I) and loci randomly isolated from genomic DNA (Type II) because the two classes of genetic markers could reveal different aspects of differentiation processes acting during the evolutionary history of the investigated species.

Table 3 Species pairwise F_{ST} and F'_{ST} values

Sample pairs	21 Loci		18 Neutral loci		9 Neutral Type II loci		9 Neutral Type I loci	
	F_{ST}	F'_{ST}	F_{ST}	F'_{ST}	F_{ST}	F'_{ST}	F_{ST}	F'_{ST}
Ch–Cm	0.2100	0.4996	0.1597	0.4066	0.1863	0.6198	0.1211	0.2501
Ch–Cr	0.1672	0.3821	0.1114	0.2919	0.0914	0.3551	0.1399	0.2765
Cm–Cr	0.2004	0.4645	0.1196	0.3030	0.1028	0.3429	0.1417	0.2893

All pairwise comparisons were performed using four datasets: all 21 loci considered in this study, the 18 neutral loci indicated by the neutrality tests (loci Cr38, Ch684, and Ch8501 were excluded), 9 neutral Type II loci (locus Cr38 was excluded), and 9 Type I loci (loci Ch684 and Ch8501 were excluded). All reported values are statistically significant (p values <0.0001). F_{ST}: actual estimate of population differentiation; F'_{ST}: standardized measure of population divergence. Ch: *Chionodraco hamatus*; Cm: *Chionodraco myersi*; Cr: *Chionodraco rastrospinosus*

Table 4 Species pairwise F_{ST} and F'_{ST} values at outlier loci indicated by neutrality tests: Cr38, Ch684, and Ch8501

Sample pairs	Cr38		Ch684		Ch8501	
	F_{ST}	F'_{ST}	F_{ST}	F'_{ST}	F_{ST}	F'_{ST}
Ch–Cm	0.3923	0.8538	0.1180	0.2730	**0.8378**	**0.9751**
Ch–Cr	**0.6338**	**0.9192**	**0.6142**	**0.8536**	0.0060	0.0065
Cm–Cr	**0.5460**	**0.9524**	**0.4324**	**0.7675**	**0.7892**	**0.9617**

All reported values are statistically significant (p values <0.0001). Pairwise comparisons in which each locus was identified as an outlier are reported in bold. F_{ST}: actual estimate of population differentiation; F'_{ST}: standardized measure of population divergence. Ch: *Chionodraco hamatus*; Cm: *Chionodraco myersi*; Cr: *Chionodraco rastrospinosus*

Table 5 Species pairwise F_{ST} and F'_{ST} 95 % confidence interval (CI) calculated at 18 putative neutral loci

Sample pairs	F_{ST} 95 % CI	F'_{ST} 95 % CI
Ch–Cm	0.0920–0.2320	0.2342–0.5906
Ch–Cr	0.0600–0.1670	0.1598–0.4375
Cm–Cr	0.0570–0.1920	0.1419–0.4890

F_{ST}: actual estimate of population differentiation; F'_{ST}: standardized measure of population divergence. Ch: *Chionodraco hamatus*, Cm: *Chionodraco myersi*, Cr: *Chionodraco rastrospinosus*

Type I loci showed a better amplification performance with no detection of null alleles and no departure from HWE (except 1 out of the 33 tests performed), which can be explained by the high conservation of primer annealing regions residing into coding sequences. On the other hand, Type II loci showed some deviations from HWE but a higher level of genetic polymorphism, as indicated by N_A, A_R, H_{Obs}, and H_{Exp} values, which significantly exceeded those calculated with Type I loci. A high level of variability can increase the power of molecular markers to find genetic differences, although this might be obscured by recurrent mutations (O'Reilly et al. 2004). F_{ST} measures are also known to be limited by the average homozygosity within populations (Hedrick 2005). In this study, F_{ST} values calculated at Type II loci exceeded those calculated at Type I loci in all pairwise comparisons between species only when they were standardized following Hedrick (2005). This finding confirms the higher power of neutral genomic loci to detect differentiation and further points out a possible bias in nonstandardized F_{ST} estimates.

Whatever combination of loci was considered, F_{ST} estimates indicated the presence of three distinct gene pools corresponding to the three nominal species, with *C. hamatus* and *C. rastrospinosus* being more closely related. This result

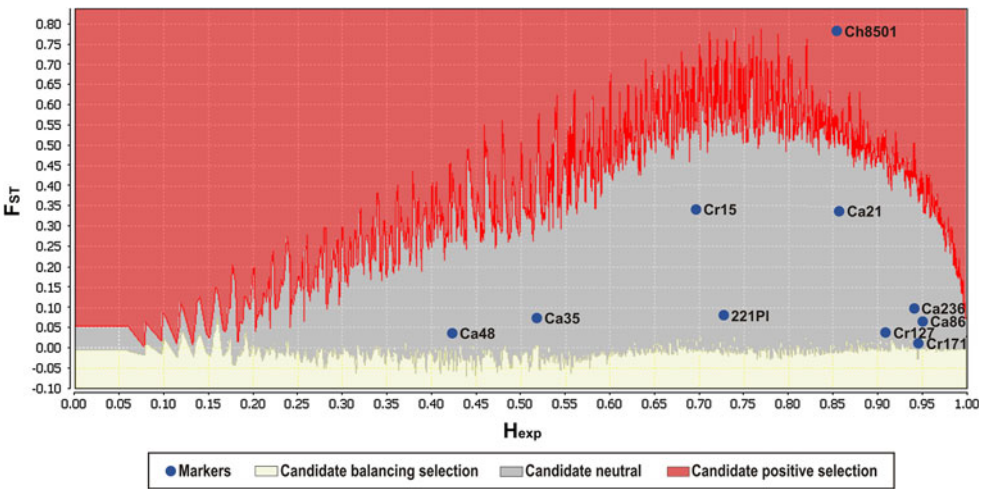

Fig. 1 Result from LOSITAN 1,000,000 simulations under the SMM for the comparison between *C. rastrospinosus* and *C. myersi* using all Type II loci, except the outlier Cr38, plus Ch8501. Shown in *gray* (center) is the 99 % CI of the expected F_{ST} versus H_{Exp} (expected heterozygosity) distribution with neutral markers. Loci outside this interval are potential candidates for being subject to directional selection (*red, top*) or balancing selection (*yellow, bottom*). (Color figure online)

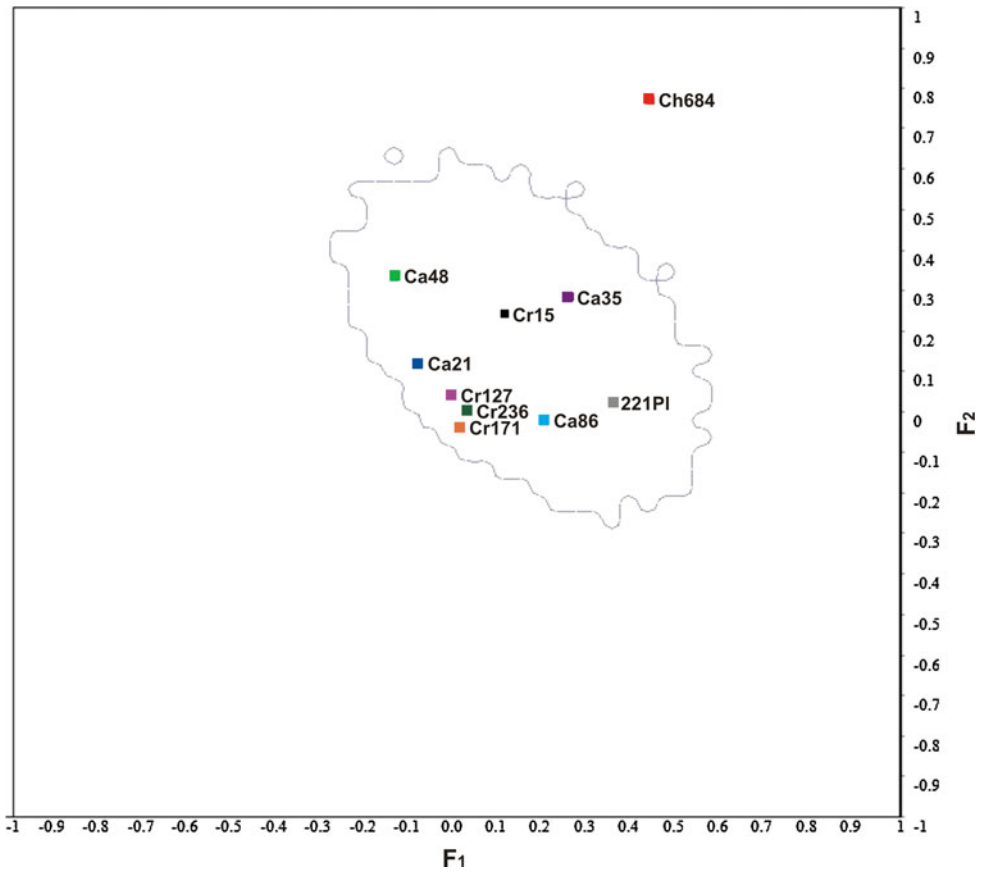

Fig. 2 Result from DETSEL 500,000 coalescent simulations under the IAM for the comparison between *C. hamatus* and *C. rastrospinosus* using all Type II loci, except the outlier Cr38, plus Ch684. The *circumscribed* area represents the 99 % probability region of the expected distribution of population-specific parameters of population divergence, F_1 and F_2, in a neutral scenario. Loci outside this area are potential candidates for being subject to directional selection

is in line with a previous mitochondrial DNA analysis showing that, despite a small genetic differentiation, all the haplotypes of *C. hamatus* and *C. rastrospinosus* clustered together in two sisters monophyletic groups, while *C. myersi* was more distantly related (Patarnello et al. 2003).

Results obtained in this study by neutrality tests suggest that, during this period of time, selective pressures might have played a role in the divergence process of the three species. Indeed, three outlier loci, displaying a higher than expected divergence, were identified: Cr38 and Ch684, in all pairwise comparisons including *C. rastrospinosus*, and Ch8501 in all pairwise comparisons including *C. myersi*. With regard to Ch684 microsatellite, BLAST search (Altschul et al. 1990) showed that it resides at the 3′ region of a gene coding for calmodulin, a highly conserved calcium-binding protein that transduces calcium signals and mediates many crucial processes in eukaryotic cells, such as inflammation, metabolism, apoptosis, muscle contraction, intracellular movement, short-term and long-term memory, nerve growth, and the immune response. BLAST similarity search of Ch8501 sequence showed that this locus is instead associated with a *C. hamatus* transcript showing high similarity to other fish mRNA sequences with no functional annotation provided. FULL-LENGTHER Web tool (Lara et al. 2007) indicated that the transcript from which Ch8501 microsatellite was identified has a high probability (92 %) to be a gene coding sequence. Cr38 microsatellite was detected as an outlier besides being a Type II marker. Type II loci were randomly isolated from genomic DNA and we cannot exclude a priori the possibility to draw by chance a microsatellite located in coding regions. In addition, several studies verified the functional relevance of a considerable number of microsatellites residing outside coding gene sequences and thus under the possible influence of natural selection (see Chistiakov et al. 2006 for review). Intriguingly, BLAST search of Cr38 sequence against the NCBI nucleotide database showed a significant alignment, with 96 % sequence identity over 92 nucleotides, to a gene encoding an AFGP/TLP in the Antarctic notothenioid *Dissostichus mawsoni* (Nicodemus-Johnson et al. 2011). Notably, genes coding for notothenioid AFGPs, which are essential molecules for freezing avoidance at subzero temperatures, evolved from a pancreatic TLP, presumably through an ancestral intermediate, that is, a chimeric AFGP/TLP gene (Chen et al. 1997; Cheng and Chen 1999).

Based on the results of neutrality tests between species, and assuming that the higher divergence of outlier loci can be attributed to directional selection, selective pressures could have operated in the branch leading to the divergence

of *C. myersi* for Ch8501 and in the branch leading to the divergence of *C. rastrospinosus* for loci Cr38 and Ch684. Recently, Near et al. (2012) showed how the diversification of Antarctic notothenioids in the Southern Ocean, which was considered a direct consequence of AFGP evolution, was related both to freezing avoidance and to the colonization and adaptation to new ecological niches created by glacial and ice sheet activity during the Late Miocene (11.6–5.3 millions of years ago), thus at least 10 million years after the origin of AFGPs. A time-calibrated Bayesian phylogeny of 83 notothenioid species, performed by the same authors, revealed pulses of lineage diversification occurring in the most derived clades of Antarctic notothenioids, such as the genus *Trematomus* and the family Channichthyidae. Therefore, selective pressures acting on specific loci, as those detected by this study, might reflect past evolutionary processes leading to local adaptation and differentiation of the three *Chionodraco* species. As mentioned before, the divergence of the *Chionodraco* genus was estimated between 2 and 1.8 millions of years ago (additional result not reported in Near et al. 2012), with the separation of *C. myersi* occurring before the divergence of *C. hamatus* and *C. rastrospinosus* (Patarnello et al. 2003). The area of distribution of the three species may give some insights into the ecological cues leading to their diversification. *C. hamatus* and *C. myersi* both display a circum-Antarctic distribution, while *C. rastrospinosus* is found in waters surrounding the Antarctic Peninsula, the South Shetland Islands, and the South Orkney Islands (Kock 1992). Patarnello et al. (2003) hypothesize that the establishment of the Scotia-Weddell confluence, where out flowing Weddell Sea waters converge with the eastward flowing waters of the Scotia Sea, may represent the vicariant event leading to interrupted of gene flow between *C. hamatus* and *C. rastrospinosus*. However, our data suggest that selective pressures might have operated in the past promoting species diversification and possibly leading to local adaptation of the *Chionodraco* species.

Acknowledgments We are grateful to Vittorio Varotto, Dr. Erica Bortolotto, the Italian National Program for Antarctic Research (PNRA), and the Alfred Wegener Institute (AWI) for collecting samples during the Antarctic expeditions. This work has been funded by the Italian National Program for Antarctic Research (PNRA). CA is a PhD student in Evolutionary Biology at the University of Padova, with a program partially funded under National Science Foundation (NSF) Grant 0741348. CP has been funded by a un Senior Research Grant of University of Padova (GRIC110B82). IAMM has been funded by a University of Padova post doc Grant (CPDR084151/08).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410. doi:10.1016/S0022-2836(05)80360-2

Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007) Automated binning of microsatellite alleles: problems and solutions. Mol Ecol Notes 7(1):10–14. doi:10.1111/j.1471-8286.2006.01560.x

André C, Larsson LC, Laikre L, Bekkevold D, Brigham J, Carvalho GR, Dahlgren TG, Hutchinson WF, Mariani S, Mudde K, Ruzzante DE, Ryman N (2010) Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): direct, simultaneous evaluation of neutral vs putatively selected loci. Heredity (Edinb) 106(2):270–280. doi:10.1038/hdy.2010.71

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. BMC Bioinformatics 9:323. doi:10.1186/1471-2105-9-323

Avise JC (1994) Molecular markers, natural history and evolution. Chapman & Hall, New York

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol 13(4):969–980. doi:10.1111/j.1365-294X.2004.02125.x

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proc R Soc Lond B Biol Sci 263(1377):1619–1626

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996–2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France)

Bowcock AM, Kidd JR, Mountain JL, Hebert JM, Carotenuto L, Kidd KK, Cavallisforza LL (1991) Drift, admixture, and selection in human-evolution—a study with DNA polymorphisms. Proc Natl Acad Sci USA 88(3):839–843. doi:10.1073/pnas.88.3.839

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol 24(3):621–631. doi:10.1093/molbev/msl191

Chen LB, DeVries AL, Cheng C-HC (1997) Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. Proc Natl Acad Sci USA 94(8):3811–3816

Cheng CHC, Chen LB (1999) Evolution of an antifreeze glycoprotein. Nature 401(6752):443–444. doi:10.1038/46721

Chistiakov DA, Hellemans B, Volckaert FAM (2006) Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. Aquaculture 255(1–4):1–29. doi:10.1016/j.aquaculture.2005.11.031

Coppe A, Agostini C, Marino IAM, Zane L, Bargelloni L, Bortoluzzi S, Patarnello T (2013) Genome evolution in the cold: antarctic Icefish muscle transcriptome reveals selective duplications increasing mitochondrial function. Genome Biol Evol. doi:10.1093/gbe/evs108

Damerau M, Matschiner M, Salzburger W, Hanel R (2012) Comparative population genetics of seven notothenioid fish species reveals high levels of gene flow along ocean currents in the southern Scotia Arc, Antarctica. Polar Biol 35(7):1073–1086. doi:10.1007/s00300-012-1155-x

Fauvelot C, Lemaire C, Planes S, Bonhomme F (2007) Inferring gene flow in coral reef fishes from different molecular markers: which loci to trust? Heredity (Edinb) 99(3):331–339. doi:10.1038/sj.hdy.6801005

Fischer W, Hureau JC (1985) FAO species identification sheets for fishery purposes. In: Fischer W, Hureau JC (eds) Southern ocean: fishing areas 48, 58, and 88, CCAMLR Convention area. Food and Agriculture Organization of the United Nations, Rome, pp 233–470

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(2):977–993. doi:10.1534/genetics.108.092221

Freamo H, O'Reilly P, Berg PR, Lien S, Boulding EG (2011) Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. Mol Ecol Resour 11(Suppl 1):254–267. doi:10.1111/j.1755-0998.2010.02952.x

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet (1995)

Hedrick PW (2005) A standardized genetic differentiation measure. Evolution 59(8):1633–1638

Kock KH (1992) Antarctic fish and fisheries. Studies in polar research. Cambridge University Press, Cambridge

Lara AJ, Pérez-Trabado G, Villalobos DP, Díaz-Moreno S, Cantón FR, Claros MG (2007) A web tool to discover full-length sequences—Full-Lengther. In: Corchado E, Corchado JM, Abraham A (eds) Innovations in hybrid intelligent systems. vol advances in soft computing. Springer, Berlín, pp 361–368. doi:10.1007/978-3-540-74972-1_47

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74(1):175–195

Matschiner M, Hanel R, Salzburger W (2009) Gene flow by larval dispersal in the Antarctic notothenioid fish *Gobionotothen gibberifrons*. Mol Ecol 18(12):2574–2587. doi:10.1111/j.1365-294X.2009.04220.x

Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution 60(11):2399–2402. doi:10.1111/j.0014-3820.2006.tb01874.x

MolecularEcologyResourcesPrimerDevelopmentConsortium, Agostini C, Agudelo PA, BÂ K, Barber PA, Bisol PM, Brouat C, Burgess TI, Calves I, Carrillo Avila M, Chow S, Cordes L, Da Silva D, Dalecky A, De Meester L, Doadrio I, Dobigny G, Duplantier JM, Evison SEF, Ford R, Fresneau D, Galetti PM, Gauthier P, Geldof S, Granjon L, GuÉRin F, Hardy GE, Hernandez Escobar C, Hima K, Hu J, Huang L, Humeau L, Jansen B, Jaquemet S, Jiang Z-Q, Jung S-J, Kim B-S, Kim C-H, Kim J-O, Lai C-H, Laroche J, Lavergne E, Lawton-Rauh A, Le Corre M, Leach MM, Lee J, Leo AE, Lichtenzveig J, Lin LIN, Linde CC, Liu S-F, Marino IAM, McKeown NJ, Nohara K, Oh M-J, Okamoto H, Oliver R, Olivera Angel M, Ornelas-GarcÍA CP, Orsini L, Ostos Alfonso H, Othman AS, Papetti C, Patarnello T, Pedraza-Lara C, Piller KR, Poteaux C, Requier JB, Roziana MK, Semba Y, Sembene M, Shah RM, Shahril AR, Shao A, Shaw PW, Song L, Souza Ferreira R, Su Y-Q, Suzuki N, Tatard C, Taylor KM, Taylor PWJ, Thiam M, Valbuena R, Wang HE, Yang B-G, Yuan Q, Zajonz U, Zane L, Zhu L, Zhuang Z-M, Zulaiha AR (2011) Permanent Genetic Resources added to Molecular Ecology Resources Database 1 October 2010–30 November 2010. Mol Ecol Resour 11(2):418–421. doi:10.1111/j.1755-0998.2010.02970.x

Near TJ, Dornburg A, Kuhn KL, Eastman JT, Pennington JN, Patarnello T, Zane L, Fernandez DA, Jones CD (2012) Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. Proc Natl Acad Sci USA 109(9):3434–3439. doi:10.1073/pnas.1115169109

Nicodemus-Johnson J, Silic S, Ghigliotti L, Pisano E, Cheng CHC (2011) Assembly of the antifreeze glycoprotein/trypsinogen-like protease genomic locus in the Antarctic toothfish *Dissostichus mawsoni* (Norman). Genomics 98(3):194–201. doi:10.1016/j.ygeno.2011.06.002

Nielsen EE, Hemmer-Hansen J, Poulsen NA, Loeschcke V, Moen T, Johansen T, Mittelholzer C, Taranger GL, Ogden R, Carvalho GR (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). BMC Evol Biol 9:276. doi:10.1186/1471-2148-9-276

O'Brien SJ, Womack JE, Lyons LA, Moore KJ, Jenkins NA, Copeland NG (1993) Anchored reference loci for comparative genome mapping in mammals. Nat Genet 3(2):103–112. doi:10.1038/ng0293-103

O'Reilly PT, Canino MF, Bailey KM, Bentzen P (2004) Inverse relationship between F and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. Mol Ecol 13(7):1799–1814. doi:10.1111/j.1365-294X.2004.02214.x

Papetti C, Zane L, Patarnello T (2006) Isolation and characterization of microsatellite loci in the icefish *Chionodraco rastrospinosus* (Perciformes, Notothenioidea, Channichthyidae). Mol Ecol Notes 6(1):207–209. doi:10.1111/j.1471-8286.2005.01194.x

Papetti C, Susana E, Patarnello T, Zane L (2009) Spatial and temporal boundaries to gene flow between *Chaenocephalus aceratus* populations at South Orkney and South Shetlands. Mar Ecol Prog Ser 376:269–281. doi:10.3354/meps07831

Papetti C, Marino IAM, Agostini C, Bisol PM, Patarnello T, Zane L (2011) Characterization of novel microsatellite markers in the Antarctic silverfish *Pleuragramma antarcticum* and cross species amplification in other Notothenioidei. Conserv Genet Resour 3(2):259–262. doi:10.1007/s12686-010-9336-9

Papetti C, Pujolar JM, Mezzavilla M, La Mesa M, Rock J, Zane L, Patarnello T (2012) Population genetic structure and gene flow patterns between populations of the Antarctic icefish *Chionodraco rastrospinosus*. J Biogeogr 39(7):1361–1372. doi:10.1111/j.1365-2699.2011.02682.x

Patarnello T, Marcato S, Zane L, Varotto V, Bargelloni L (2003) Phylogeography of the *Chionodraco* genus (Perciformes, Channichthydae) in the Southern Ocean. Mol Phylogenet Evol 28(3):420–429. doi:10.1016/s1055-7903(03)00124-6

Patwary MU, Kenchington EL, Bird CJ, Zouros E (1994) The use of random amplified polymorphic DNA markers in genetic-studies of the sea-scallop *Placopecten magellanicus* (Gmelin, 1791). J Shellfish Res 13(2):547–553

Porter AH (2003) A test for deviation from island-model population structure. Mol Ecol 12(4):903–915. doi:10.1046/j.1365-294X.2003.01783.x

Raymond M, Rousset F (1995) GENEPOP (Version-1.2)—population-genetics software for exact tests and ecumenicism. J Hered 86(3):248–249

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43(1):223–225

Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour 8(1):103–106. doi:10.1111/j.1471-8286.2007.01931.x

Schlötterer C (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics 160(2):753–763

Seeb LW, Templin WD, Sato S, Abe S, Warheit K, Park JY, Seeb JE (2011) Single nucleotide polymorphisms across a species' range: implications for conservation studies of Pacific salmon. Mol Ecol Resour 11(Suppl 1):195–217. doi:10.1111/j.1755-0998.2010.02966.x

Susana E, Papetti C, Barbisan F, Bortolotto E, Buccoli S, Patarnello T, Zane L (2007) Isolation and characterization of eight microsatellite loci in the icefish *Chaenocephalus aceratus* (Perciformes, Notothenioidei, Channichthyidae). Mol Ecol Notes 7(5):791–793. doi:10.1111/j.1471-8286.2007.01703.x

Van de Putte AP, Van Houdt JKJ, Maes GE, Janko K, Koubbi P, Rock J, Volckaert FAM (2009) Species identification in the trematomid family using nuclear genetic markers. Polar Biol 32(12):1731–1741. doi:10.1007/s00300-009-0672-8

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting

genotyping errors in microsatellite data. Mol Ecol Notes 4(3):535–538. doi:10.1111/j.1471-8286.2004.00684.x

Vasemagi A, Nilsson J, Primmer CR (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). Mol Biol Evol 22(4):1067–1076. doi:10.1093/molbev/msi093

Vitalis R (2003) DetSel 1.0: a computer program to detect markers responding to selection. J Hered 94(5):429–431. doi:10.1093/jhered/esg083

Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. Genetics 158(4):1811–1823

Ward RD, Woodwark M, Skibinski DOF (1994) A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. J Fish Biol 44(2):213–232. doi:10.1111/j.1095-8649.1994.tb01200.x

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38(6):1358–1370