

The original publication is available at

<http://www.nrcresearchpress.com/>

Genetic connectivity of the reef fish *Lutjanus kasmira* at the scale of the western Indian Ocean

D. Muths^{a,*}, G. Gouws^b, M. Mwale^b, E. Tessier^c, J. Bourjea^a

^a IFREMER - Délégation de la Réunion, Le Port, Réunion, France

^b Réserve Naturelle Marine de La Réunion, La Saline les Bains, Réunion, France

^c South African Institute for Aquatic Biodiversity, Grahamstown, South Africa

*: Corresponding author : D. Muths, email address : muthsdelphine@yahoo.fr

Résumé :

Étudier la structure génétique des espèces permet d'estimer le degré de connectivité évolutive qui existe entre différentes localités pour ces espèces; cette connaissance est importante dans le cadre de la protection de la biodiversité marine, et plus particulièrement pour la délimitation d'aires marines protégées. Dans ce contexte, 601 *Lutjanus kasmira* (Forsskål, 1775) ont été échantillonnés dans 16 localités de l'ouest de l'Océan Indien. La structure génétique existant entre ces localités a été déterminée par le séquençage du gène mitochondrial cytochrome *b* et le génotypage des individus à huit loci microsatellites. Les deux marqueurs génétiques indiquent une absence de différence significative entre les sites échantillonnés, et ce même entre des échantillons distants de plus de 4000 km. Cette absence de structure génétique est principalement liée à l'écologie de l'espèce, d'une part son importante plasticité écologique et d'autre part sa forte capacité de dispersion. Néanmoins, de légères différences génétiques sont observées pour les sites de Maurice et Moroni, ainsi qu'une structure entre individus au sein de chaque localité et restent encore à expliquer notamment par l'étude des processus locaux de renouvellement des populations.

Abstract :

Examining the genetic structure of species allows an estimate of the level of evolutionary connectivity between localities; this information is important for marine biodiversity protection, in particular, for the delineation of marine protected areas. In this context, a total of 601 *Lutjanus kasmira* (Forsskål, 1775) were sampled in 16 localities of the western Indian Ocean and analyzed with both mitochondrial cytochrome *b* sequencing and eight microsatellite loci genotyping. Both genetic markers indicate that differentiation was not significant even between samples separated by more than 4000 km. This absence of genetic differentiation among samples was favored by ecological plasticity of the species and is now ensured by resultant high levels of dispersal. Nevertheless, some significant genetic structure was detected for the areas of Mauritius and Moroni, as well as within populations in all localities, which will have to be explained by additional studies on local processes.

1. Introduction

Examining the genetic structure of species allows for estimates of the level of individual exchange between populations and the identification of barriers to gene flow. This constitutes a recognized indirect approach to delineate population boundaries (Palumbi 2003) and thus to assist in the elaboration of the most efficient strategies in preserving marine biodiversity. Indeed, even when populations are demographically self-sustaining they need to be linked to each other to promote recovery from local extinctions (Allison et al. 1998, Allison et al. 2003). Thus, integrating information on evolutionary connectivity between populations in the design of marine protected areas network increase the global efficiency of biodiversity protection. An interesting application of genetic tools in a management context was the comparison of genetic structure of 27 marine species in the Hawaiian Archipelago (Toonen et al. 2010) and the consequent suggestion that the most effective ecosystem-based management approach required at least five spatially managed regions. One study based on 50 rocky intertidal invertebrate species on the Northwestern American coast indicated that an important barrier to marine dispersal had been underestimated in MPAs design (Kelly and Palumbi 2010). Such breaks may disrupt the connections among protected areas, and greatly reduce the network's ability to stabilize marine diversity.

Existing marine reserves in the Western Indian Ocean (WIO) appear largely ineffective and remain insufficient as a whole for the protection of marine reef diversity (Mora et al. 2006). Concerning the Western Indian Ocean – defined here as the waters bounded by eastern coast of Africa and as far East as 75° E -, the first step in this process might be the acquisition of population genetic data. As recently emphasized (Ridgway and Sampayo 2005, Gaither et al. 2010), very few genetic studies are available in the WIO and it is impossible to draw general conclusions about patterns of connectivity of marine life. Even if most of the studies indicate high levels of connectivity between localities of the WIO (Ridgway et al. 2001, Silva et al. 2010), they also notice that specific areas might be considered as independent management units (MUs); such as the example of green turtle *Chelonia mydas* with individuals nesting at the South and at the North of the Mozambique Channel belonging to separate genetic stocks (Bourjea et al. 2007). Studies on the parrot fish *Scarus ghobban* (Visram et al. 2010) and on the mangrove crab *Neosarmatium meinerti* (Ragionieri et al. 2010) indicate the existence of genetic isolation between Seychelles and other sites of the south-western Indian Ocean. Each of these studies provides some crucial information about connectivity in the WIO, but one shortcoming is that these studies are only based on a single genetic marker type (most often a fragment of the mitochondrial DNA sequence). The drawbacks of employing one mtDNA gene fragment alone in genetics have long been recognized, mainly because it constitutes a single information revealing only a fraction of the evolutionary history of a species (Avice 1994). Even if the uniparental inheritance of mtDNA tends to accentuate genetic differences among population compared to nuclear genes, it does not capture the entire genetic history that is fundamental in the case of defining population structure (Paul 2000). A recent genetic study on *Myripristis berndti* (Muths et al. 2011) indicated restricted connectivity in the SWIO on the basis of microsatellite, whereas this species was assumed to be widely dispersing on the basis of mtDNA (Craig et al. 2007). Moreover, some genetic studies showed discrepancies when using several genetic markers that could have various origins: sex-biased dispersal (Keeney et al. 2005), hybridization and introgression (Arnold 1993) or population size changes (Larmuseau et al. 2010), among others. The example of the marine goby *Pomatoschistus minutus* (Larmuseau et al. 2010) showed large differences in the degree of population differentiation in Europe between the nuclear and mitochondrial markers (at least 30 times higher with mtDNA) that might mainly be explained by a recent demographic expansion. Such mitochondrial-nuclear discrepancies could have important consequences for interpretation and implications in terms of management (Monsen and Blouin 2003). These studies highlight the strong limitations of basing population delineation on one marker type alone and the obvious advantages of using combined molecular approaches, especially when such studies have concrete conservation implications such as MUs delineation.

As the present study aims to contribute to WIO-MPAs network design by increasing the information on evolutionary connectivity in this area, we worked with both mitochondrial

cytochrome *b* sequences and nuclear markers (8 microsatellite loci) datasets to be as precise as possible, for the reef fish *Lutjanus kasmira* (Forsskal, 1775). *L. kasmira* is a widespread Indo-Pacific fish species, with a natural range from South Africa to the Central Pacific (Allen and Talbot 1985). *L. kasmira* inhabits a wide depth range from shallow water to at least 265m (Allen and Talbot 1985). It is a broadcast spawner with high fecundity (Grimes 1987). We sampled and analysed 601 individuals of *L. kasmira* from 16 sites in the WIO (from Europa island in the south-west of the study area to Maldives in the north-east). The main objective is to determine the level of connectivity within this species in the WIO and to highlight whether patterns are congruent between the two genetic markers.

2. Material and Methods

2.1. Sampling sites

Sixteen sites have been sampled in the WIO between 2008 and 2010: eight in the Mozambique Channel, two on the East coast of Africa, one in Seychelles, four in the East of Madagascar and one in the Maldives (Table 1; Figure 1). Distances between sites range from 80 km between Moheli and Moroni (Comoros archipelago) to more than 4000 km between localities within the Mozambique Channel and Maldives.

Samples of *L. kasmira* were obtained by fishing, and when possible, we tried to sample fishes from the same place and size cohort to maximize sampling homogeneity (Table 1). Tissues were fixed in 95% ethanol and stored at -20°C .

2.2. Genetic analyses

Total genomic DNA was extracted using DNAeasy Tissue Kit (Qiagen) following the manufacturer's instructions. A 727 base pair (bp) fragment of the mitochondrial cytochrome *b* region was amplified by PCR using the primers CytbH (Song et al. 1998) and CytbR (Taberlet et al. 1992). Reactions were performed in 20 μl containing 1X PCR buffer, 2 mM MgCl_2 , 20 μM of each dNTPs, 0.5 μM of each primer, 0.5 U of Silverstar DNA Polymerase Taq (Eurogentec), 25 ng of genomic DNA. Cycling parameters were 94°C for 3 min, followed by 35 cycles of 94°C for 45 s, 54°C for 45 s, and 72°C for 45 s and a final extension at 72°C for 4 min. PCR products were purified and sequenced on an ABI 3100 sequencer (Applied Biosystems Inc.). Sequences were generated in forward and reverse directions. They were checked in BioEdit Sequence Alignment Editor (Hall 1999) and aligned using ClustalW (Thompson et al. 1994). Cytochrome *b* sequences of this species were available (accession numbers FJ754049-FJ754133; Gaither et al. 2010) but are shorter (475bp); hence, all the present haplotypes were submitted to GenBank (accession numbers JF514414-JF514500).

Eight of the thirteen microsatellite loci previously developed for this species (Lk-08, Lk-10, Lk-18, Lk-27, Lk-29, Lk-30, Lk-49, Lk-57) were used following the recommended conditions (Molecular Ecology Resources Primer Development Consortium et al. 2011); the five remaining loci were avoided due to heterozygote deficiencies, significant presence of null alleles and discovery of short allele dominance as detected by Micro-Checker (Van Oosterhout et al. 2004). Amplified fragments of the eight microsatellite loci were separated on an ABI Prism 3100 genetic analyzer. Alleles were scored using a co-migrating size standard (Genescan500, Applied Biosystems Inc.) and identified using GeneMapper4 (Applied Biosystems Inc.).

2.3. Data analyses

For cytochrome *b* sequences, haplotype (*h*) and nucleotide (π) diversities and Tajima's (1989) D-statistic were estimated per locality with DNAsp 5.0 (Librado and Rozas 2009). Tajima's D-statistic tests for departure from equilibrium between the addition of variation by mutation and the removal of variation by genetic drift; theoretically, mutation-drift equilibrium should be reached if the effective population size has remained stable in the past. A mismatch curve was examined with the same software package. Mismatch distribution parameters τ , θ_0 and θ_1 were estimated with Arlequin 3.5 (Excoffier and Lischer 2010), from which demographic parameters could be estimated as in Gaither *et al.* (2010): they considered a generation time of 3.7 years and a sequence divergence estimate of 1–2% per Myr between lineages. The age of each population was calculated in years using the equation $\tau = 2 \text{ u.t.}$, where *t* is the age in generations and *u* is the mutation rate per generation. As underline by these authors (Gaither *et al.* 2010), and due to many approximations, these values have to be taken as relative values. Effective female population size (N_{ef}) was calculated for time zero and present day using the equation $\theta = 2N_{ef} \cdot u$. Pairwise values of genetic differentiation were estimated with the fixation index for sequence data Φ_{st} as executed by Arlequin 3.5 (Excoffier and Lischer 2010) and tested using 10000 permutations. Critical significance levels for multiple testing were corrected by calculating the false discovery rate (FDR) using the program Q-value (Storey and Tibshirani 2003). This method estimates the number of false positives within the set of significant values ($p < 0.05$) and then calculates a new significance probability, *q*. The test is considered to be significant, if $q < 0.05$. Jost's (2008) unbiased estimator of divergence (*D*, based on the effective number of alleles rather than on the expected heterozygosity as Φ_{st}) was also calculated per pair of localities using SPADE (available at <http://chao.stat.nthu.edu.tw/softwareCE.html>). DNAsp 5.0 (Librado and Rozas 2009) was used to estimate the nearest-neighbour statistic *Snn* (Hudson 2000). *Snn* is a measure of how often the 'nearest neighbours' in sequence space are from the same locality in geographical space. *Snn* varies from 0 to 1: under 0.5, it is assumed that populations are in panmixia and values closer to 1 indicate that populations are differentiated. Median-joining network (Bandelt *et al.* 1999) was constructed using Network 4.5 (available at <http://www.fluxus-technology.com/>).

For microsatellites, allele frequencies, mean number of alleles per population (*Nall*), and the observed (*Ho*) and expected (*He*) heterozygosities (Nei 1987) were calculated with Arlequin 3.5 (Excoffier and Lischer 2010). Allelic diversity was adjusted to account for differences in sample size by estimating the allelic richness (*Rs*) by rarefaction process using Fstat 2.9.3.2 (Goudet 1995). Deviations from Hardy-Weinberg equilibrium were examined for each population, at each locus, by calculating Wright's (1969) fixation index F_{is} as estimated by Weir and Cockerham (1984) and tested using exact tests performed with Arlequin 3.5 (Excoffier and Lischer 2010). Genetic differentiation between populations were estimated by calculating Wright's F_{ST} statistic (Weir and Cockerham 1984) and then tested using exact tests (10000 permutations) for the null hypothesis of identity of allelic distributions across populations. Critical significance levels for multiple testing were also corrected using the program Q-value (Storey and Tibshirani 2003). Jost's (2008) unbiased estimator of divergence (*D*) was also calculated per pair of localities using SMOGD (Crawford 2009). To determine if the genotypes belonged to one or more genetic pools, microsatellite data were analysed using the software STRUCTURE 2.3.2 (Pritchard *et al.* 2000) which uses an iterative computation process to infer the most likely number of populations (*K*) represented in the total sample. For this analysis, an admixture model assuming independent allele frequencies was used and twenty replicates were run (each with $1 \cdot 10^5$ burn-in samples/generations and $5 \cdot 10^5$ iterations) for *K* values from 1 to 16.

For both mitochondrial sequences and microsatellite data, the geographic partitioning of genetic structure was investigated in different ways. Firstly, marine distances between localities were estimated on the <http://www.geodistance.com> website and plotted against genetic distance to test

for patterns of isolation-by-distance based on a linear relationship following the recommendations of Rousset (1997), using $\Phi_{st} / (1 - \Phi_{st})$ for mitochondrial data or $F_{ST} / (1 - F_{ST})$ for microsatellite data. The significance of this relationship was tested with a Mantel test, performed in R (R Development Core Team 2010) using the *ncf* package (available at <http://onb.ent.psu.edu/onb1/R>). Then, spatial genetic variation was assessed via spatial autocorrelation analysis as implemented in GenAlEx 6 (Peakall and Smouse 2006). The significance of the spatial autocorrelation coefficient (r) between geographic and genetic distances was determined by generating bootstrap 95% confidence intervals for distance classes from 50 to 1000 km. Then, Arlequin 3.5 (Excoffier and Lischer 2010) was used to perform analysis of molecular variance (AMOVA) for both marker sets, independently, with *a priori* groupings based on geographical proximity and oceanic current (localities on the East side of Madagascar versus localities on the West (that could be subdivided between localities within the Mozambique Channel and the northern localities); see Figure 1 for currents and Table 2 for grouping details). Finally, the software SAMOVA 1.2 (Dupanloup et al. 2002) was used to perform spatial analysis of molecular variance (SAMOVA). This approach detects genetic barriers in a sampling region without *a priori* definition of groups and identifies geographic partitions that maximize genetic differences between groups and geographic homogeneity within groups; it was tested for K group values ranging from 1 to 4, with 100 annealing replicates each time.

3. Results

A total of 601 individuals of *Lutjanus kasmira* were sampled in 16 localities of the West Indian Ocean: all were analysed with microsatellites, while 576 were sequenced at the cytochrome *b* locus.

3.1. MtDNA marker

The 727-bp mitochondrial cytochrome *b* sequences revealed a high level of nucleotide and haplotype polymorphism in *L. kasmira*. Sequence analysis indicated the occurrence of 74 polymorphic sites, from which 87 different haplotypes were identified. The haplotype diversity (h) and mean nucleotide diversity (π) of the whole dataset were 0.767 and 0.0024, respectively. Genetic diversities within each locality are provided in Table 3. The network constructed with the WIO haplotypes (Figure 2) revealed a 'star-like' pattern with the most common haplotypes at the centre, from which a crown of less common haplotypes radiated. The most common haplotype (representing 47% of all 576 sequences) was well represented in the 16 localities with a frequency varying from 25% in Seychelles to 62% in Tanzania (Table 3, Figure 2). Consequently, the haplotype diversity was lowest in Tanzania ($h = 0.623$) and highest for Seychelles ($h = 0.941$). The most common haplotype was generally more common in the localities of the Mozambican Channel [52% and greater, with the exception of Nosy Be (42%)] than in the others [43% and less, with the exception of Tanzania]. The proportion of private haplotypes (Table 3) ranged from 0 % in Tanzania and Maldives to 35 % at Madagascar, with a mean of 9 %. Tajima's D values were negative and significant for the entire sample ($D = -2.20$, $p < 0.001$) and for the individual localities, except for Mauritius. These negative values indicated a significant excess of low frequency variants, indicating a recent coalescence time for the WIO *L. kasmira* populations, while Mauritius showed the signature of a more stable population. A mismatch distribution was constructed using the 576 sequences and shows a unique peak with a low and insignificant Harpending's raggedness index ($r = 0.01$, $p = 0.81$) indicating no departure between observed and expected unimodal distributions. The population age estimated for the whole samples is 144 000 - 289 000 years with an initial effective population of 2 to 4 thousands females and a current effective population of 400 to 900 millions females. Demographic parameters estimated per locality are provided in Table 3.

Pairwise values of genetic differentiation (Φ_{ST}) are provided in Table 4. Overall Φ_{ST} was 0.008 ($p < 0.001$). Of the 120 comparisons, 110 were very low and non-significant ($\Phi_{ST} < 0.055$). Consistent with this large majority of low differentiation values, values of Jost's D were zero for most localities comparison (Table 4). Of the 10 significant comparisons ($0.022 < \Phi_{ST} < 0.052$; $p < 0.05$), eight comparisons involved Mauritius and three the Seychelles localities. Considering Q-values, only two comparisons were still significantly different: Mauritius *versus* Glorieuses or *versus* Juan de Nova. Consistent with these significant differentiation results, values of Jost's D (Table 4) were the highest when comparisons included Seychelles or Mauritius. The highest value was for the comparison between Seychelles and Mauritius ($D = 0.233$). The Mauritius differentiation could be explained by two secondary haplotypes which were common in this locality (17.4% each; see Figure 2), whereas these were very rare elsewhere (less than 3%). All the other localities were characterised by less common (less than 5% in mean) and more evenly shared secondary haplotypes. The differentiation of Seychelles might be due to the lowest frequency of the main haplotype and the high haplotype diversity.

To further test genetic homogeneity, the nearest-neighbour statistic was calculated and revealed no association between sequence similarity and geographical location ($Snn = 0.092$). No isolation-by-distance pattern was identified ($r = -0.009$, $p = 0.41$). No significant autocorrelation (r) between geographic and mitochondrial distances was found regardless of the distances separating the chosen localities (see Figure 3a, for 500km results). However, AMOVA (Table 2) identified some geographic structure: if grouping was done in agreement with geographic regions, some significant variance was associated among the groups. Thus, grouping all the localities to the west of Madagascar *versus* those to the east induced a significant result ($\Phi_{CT} = 0.012$; $p < 0.01$), as well as when localities to the north were considered as a third group ($\Phi_{CT} = 0.010$; $p < 0.001$). One explanation for these significant groupings is the lower haplotype diversity and the higher proportion of the most-common haplotype observed in most localities of the Mozambican Channel. The SAMOVA did not identify this structure. Thus, without *a priori* geographic grouping, between-group variance was maximised (6% of genetic variance explained, $p = 0.04$) when two groups were considered with only one divergent locality – Mauritius – isolated from the other localities.

3.2. Microsatellite markers

The 8 loci analyzed were found to be polymorphic, with a total of 124 alleles detected in the 601 samples. No loci were in disequilibrium ($p < 0.001$) over the whole dataset, supporting the independent assortment of alleles at different loci. Allelic richness was of the same order between the 16 localities, with a mean value of 6.17 ± 0.36 (Table 5). Exclusive alleles (*i.e.* when an allele at a given locus was found exclusively in one locality) were very rare (less than one per population). The mean observed and expected heterozygosities across populations were 0.441 and 0.595, respectively. Heterozygote deficiencies were highly significant in most localities ($p < 0.001$), with Fis values ranging from 0.039 for Seychelles to 0.217 for Juan de Nova (Table 5). Heterozygote deficiencies were not locus- or population-dependant (see Supplementary Materials section).

Pairwise multilocus F_{ST} values of microsatellite differentiation are provided in Table 6. Overall F_{ST} was 0.007 ($p < 0.001$). Of the 120 comparisons, 24 were significant ($0.009 < F_{ST} < 0.021$; $p < 0.05$). Most of the significant values concern comparisons involving Moroni or Mauritius samples. Considering exact tests and Q-values, all the comparisons became non-significant. Similarly, values of Jost's D were very low (Table 2). The highest value was obtained for Moroni *versus* Mayotte. No relationship between genetic and geographic distances was identified ($r = -0.11$, $p = 0.21$) in the test for isolation-by-distance. A significant correlation ($r = 0.25$, $p < 0.05$) existed between the two genetic (mitochondrial and microsatellite) distances. Spatial autocorrelation did not differ significantly from zero (see Figure 3b for 500 km classes) whatever the distances separating localities.

The STRUCTURE analysis suggested that the highest likelihood of obtaining the present data was to consider that only one single genetic group existed ($K = 1$). For any $K > 1$, the individuals'

posterior assignment probability was essentially the same for each specified cluster, providing evidence against detectable population subdivision.

The two AMOVAs (Table 2) with same geographic grouping as for mtDNA data showed that 99% of the variance was associated at the within-locality level ($\Phi_{sc} = 0.07$; $p < 0.001$) and that the variance associated among localities within groups was also significant, but that no significant variance was associated with the partitioned groups. The SAMOVA analysis also failed to identify any significant between-group structure (less than 1% of genetic variance, $p > 0.05$).

4. Discussion

Global connectivity in the Western Indian Ocean

This study aimed to evaluate evolutionary connectivity of *Lutjanus kasmira* sampled from 16 different localities of the Western Indian Ocean. Analyses of cytochrome *b* sequences and microsatellite polymorphisms both indicated a high level of genetic homogeneity within the studied area. Mitochondrial sequences of *L. kasmira* revealed high haplotype diversity within each of the sixteen sampling sites. Most of these haplotypes were shared equally among the localities. Consequently, most of the pairwise Φ_{ST} values between localities were very low and not significantly different from zero. This together with non-significant *Snn* nearest-neighbour statistics, the absence of an isolation-by-distance pattern and spatial autocorrelation, all indicate a lack of clear mitochondrial structure within the WIO. Similarly, all microsatellite analyses indicated a high level of connectivity between *L. kasmira* localities. The microsatellite dataset was characterized by very low and non-significant pairwise F_{ST} and Jost's *D* values, with no relationship existing between geographic and genetic distances. Neither was any clear genetic structure recognized by the method implemented in STRUCTURE. All these results lead us to conclude that there is a high level of genetic homogeneity for *L. kasmira* among the different sampled localities of the WIO. Although the sampling was done to ensure a high spatial coverage; the observed gene flow among populations seems to encompass the whole study area (> 4000 km). This absence of strong genetic structure is in agreement with a recent phylogeographic study (Gaither et al. 2010) that revealed no difference at the Indo-Pacific scale for this species. Based on mitochondrial sequences and nuclear introns, this study suggested high within-species connectivity, with the only exception being the Marquesas (East Pacific) which appeared strongly isolated from all other populations. This is contrary to patterns seen in an other reef fish *Myripristis berndti* which showed overarching patterns of widescale connectivity across the Indo-Pacific (Craig et al. 2007), but with evidence of regional differentiation with the inclusion of more localities with a finer-scale focus on the SWIO (Muths et al. 2011).

Some ecological characteristics of *L. kasmira* probably influence and contribute to genetic homogeneity. The fact that this species is a good competitor for space (Schumacher and Parrish 2005), has a broad depth preference (2-265 m; Allen and Talbot 1985) and opportunistic trophic behaviour (Oda and Parrish 1987) presents several ecological elements that seem to be associated with species showing lower level of genetic structure. In a comparative study on five species of Atlantic wrasses, Rocha *et al.* (2005) showed that species with broad habitat preferences presented weak genetic population structure than specialised species. In fact, it seems that ecological specialisation affects historical demography and thus impacts population genetic structuring. Thus, generalist species are supposed to be less impacted by glacial events and thus show less structure. Lagoon specialist species in French Polynesia were thus found to have experienced stronger genetic bottlenecks than species inhabiting the outer reef slope during glacial events and thus presented more structure (Fauvelot et al. 2003). Similarly, a comparative study between two butterflyfish across the Pacific Ocean showed that the dietary specialist had undergone genetic bottlenecks while no such evidence was found for the dietary generalist species (Lawton et al. 2011). Results obtained on *L. kasmira* appear congruent with such an hypothesis with the haplotype network 'star-like' pattern and the unimodal mismatch curve being arguments against recurrent bottlenecks. Thus the opportunistic and generalist characteristics of

the species help with understanding how no major population differentiation was observed in *L. kasmira*.

The pelagic larval phase is the agent of actual connectivity between populations for most reef fishes and, thus helps to maintain genetic homogeneity within a species. The PLD of *L. kasmira* - around 30 days (Pothin 2005) - could allow larvae to passively cover distances of several hundreds of kilometers, in an area where marine currents are of 20-30 Sverdrup (Schott et al. 2009). In the present case, the pronounced swimming capability of Lutjanidae larvae (Fisher et al. 2005) might enhance dispersal ability. This study showed that Lutjanidae larvae could swim two to four times faster than the average current speed, suggesting that Lutjanidae species are able to influence dispersal distances and their spatio-temporal patterns of settlement. This large potential for dispersion and the consequent absence of genetic structure have been shown for another snapper, *L. fulviflamma* (Dorenbosch et al. 2006) in the same area. In addition, *L. kasmira* shows a large estimated effective population size, that is coupled with a high fecundity (Grimes 1987) and spawning occurring all year long along East African coast (Nzioka 1979). It was observed that the main event of recruitment occurred in austral summer (Nzioka 1979) with a variability of three months observed in Reunion between 2008 and 2011 (M. Pinault, pers. com.). These factors indicate a level of flexibility in reproductive strategy that may ensure that larvae are placed in a favorable environment, to increase the chance of settlers to survive. Improved settler survival transforms potential connectivity into adequate contribution to the adult population and finally maintains the genetic homogeneity within the species. Therefore both ecological and reproductive characteristics of *L. kasmira* help to explain why no major population differentiation was identified at the scale of the WIO.

Some restricted differentiation

The few studies previously conducted in the WIO mostly demonstrate high levels of marine connectivity between localities (Ridgway et al. 2001, Ragionieri et al. 2010, Visram et al. 2010). However, in every case, genetic differences isolate some peripheral area: for both the parrot fish *Scarus ghobban* (Visram et al. 2010) and the mangrove crab *Neosarmatium meinerti* (Ragionieri et al. 2010), the Seychelles appeared phylogeographically isolated from East African localities. In the case of the coral *Pocillopora verrucosa* (Ridgway et al. 2001), Southern Mozambique was identified as a separate cluster from more southern localities of the African coastline. Some geographically restricted differences also appeared in the present study, but, unlike these previous studies, these were not at the edge of the sampling area. On the basis of microsatellites, Moroni appears as the most divergent locality. On the basis of mitochondrial sequences, Mauritius appeared significantly different from most localities of the Mozambique Channel. The level of differentiation between Moroni or Mauritius and other localities is higher than differentiation found between Maldives - the most distant locality - or Rodrigues - the most eastern one - and the other WIO localities. These particular structures should thus be more the signature of local processes rather than due to permanent geographic isolation.

Finally, significant departure from Hardy-Weinberg equilibrium was observed in most localities. Even if some problems of allele scoring and polymorphism detection persist and could be part of the explanation, the fact that *F_{is}* values are not locus-dependant induces that an alternative hypothesis (such as temporal or spatial population substructure) is needed to explain the within-population differentiation observed with microsatellites. We are presently unable to determine the reasons for such within-population differentiation in *L. kasmira*, but recent studies offer interesting perspectives. Small-scale genetic patchiness has been firstly reported (Johnson and Black 1982) and since widely discussed (Hedgecock 1994, David et al. 1997) in invertebrate species that show substantial larval dispersal capability and large-scale genetic homogeneity. Recently, this has been reported in a reef fish *Stegastes partitus* (Christie et al. 2010, Hogan et al. 2010). These two studies invoked different hypotheses to explain such substructure of populations in two different reef systems. If different origins of settling larvae due to stochastic recruitment in the sea might best explain the genetic patchiness encountered in Meso-American barrier reef system (Hogan et al. 2010), the 'sweepstakes hypothesis' (high variance in reproductive success, implying low effective number of progenitors at the origin of a cohort) is favored in the Bahamas system

(Christie et al. 2010). Thus, even in the same species, the two phenomena could coexist, depending on the current systems in which fish are living. Both explanations were also advanced to explain the patchiness encountered in *Paralabrax clathratus* off the coast of California (Selkoe et al. 2006). An interesting perspective in the case of *L. kasmira* will be provided with access to larvae and recruits. For example, a study of *L. campechanus* (Saillant et al. 2010) indicated, that despite genetic homogeneity within the species, genetic structure occurred between young-of-the-year at a small geographical scale. Determining the scale and the importance of local processes have helped to define the metapopulation organization and the consequent management of the species (Saillant et al. 2010).

This large-scale genetic study was the first important step in understanding the evolutionary connectivity of *L. kasmira* in the WIO and leaves interesting perspectives. From a biodiversity protection perspective, the high connectivity observed supports the idea of a coordinated management for all localities in the WIO for *L. kasmira*. But single-species studies are clearly insufficient as a tool for managers in a context of global biodiversity protection. An effort is now needed to collect additional data in the WIO – from other fish species with contrasted history traits of life, but also from algae and invertebrates - to be compared and integrated in the context of MPA delineation.

Acknowledgements

We gratefully acknowledge the authorities of the different countries for permitting our sample collection in their water, in particular when it was in Marine Protected Areas: TAAF, DAF de Mayotte, Seychelles Fishing authorities and the Marine Park of Mohéli. We are very grateful to all people who helped us in the collection of samples, mainly: SHOALS of Rodrigues, Antsiva board, H. Evano, L. Le Rû, H. Sauvignet, R. Bochard, R. Le Goff, J. Wickel and F. Beudard. Samples from Maldives have been given by L. Vigliola and L. Bigot, thanks to their project IFB (Institut Français de la Biodiversité) on Maldives. A special thanks to L. Gagnevin and K. Vital from CIRAD-3P for their help at the genotyping platform. A tremendous thank you to three anonymous reviewer for their constructive remarks on the first version of the manuscript and to M. Pinault for interesting discussion on the *L. kasmira* population biology in Reunion island. This work was mainly financed by European Funds - the European Cooperation Funds EU POCT-OI 2007-2013 (project CAMP) and the EU 7th Funding Program - Capacities, Research Potential (project RUN Sea Sciences) -, Région Réunion, DEAL de La Réunion, and a Marine Sciences for Management (MASMA) grant.

Bibliography

- Allen, G., and Talbot, F. 1985. Review of the snappers of the genus *Lutjanus* (Pisces: Lutjanidae) from the Indo-Pacific, with the descriptions of a new species. *Indo-Pacific Fishes* 11: 1-87.
- Allison, G.W., Gaines, S.D., Lubchenco, J., and Possingham, H.P. 2003. Ensuring Persistence of Marine Reserves: Catastrophes Require Adopting an Insurance Factor. *Ecol Appl* 13(1): S8-S24. doi: 10.2307/3099994
- Allison, G.W., Lubchenco, J., and Carr, M.H. 1998. Marine reserves are necessary but not sufficient for marine conservation. *Ecol Appl (Supplement)* 8(1): 79-92.
- Arnold, J. 1993. Cytonuclear disequilibria in hybrid zones. *Annu. Rev. Ecol. Syst.* 24: 521-554.
- Avise, J.C. 1994. *Molecular markers, natural history and evolution*. Springer, New York.
- Bandelt, H.J., Forster, P., and Röhl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biological Evolution* 16: 37-48.

- Bourjea, J., Lapegue, S., Gagnevin, L., Broderick, D., Mortimer, J.A., Ciccione, S., Roos, D., Taquet, C., and Grizel, H. 2007. Phylogeography of the green turtle, *Chelonia mydas*, in the Southwest Indian Ocean. *Mol Ecol* 16(1): 175-186. doi: 10.1111/j.1365-294X.2006.03122.x
- Christie, M.R., Johnson, D.W., Stallings, C.D., and Hixon, M.A. 2010. Self-recruitment and sweepstakes reproduction amid extensive gene flow in a coral-reef fish. *Mol Ecol* 19(5): 1042-1057. doi: 10.1111/j.1365-294X.2010.04524.x
- Craig, T., Eble, J.A., Bowen, B.W., and Robertson, D.R. 2007. High genetic connectivity across the Indian and Pacific Oceans in the reef fish *Myripristis berndti* (Holocentridae). *Mar Ecol Prog Ser* 334: 245-254.
- Crawford, N.G. 2009. SMOGD: software for the measurement of genetic diversity. *Mol Ecol Res* 10(3): 556-557. doi: 10.1111/j.1755-0998.2009.02801.x
- David, P., Berthou, P., Noel, P., and Jarne, P. 1997. Patchy recruitment patterns in marine invertebrates: a spatial test of the density-dependent hypothesis in the bivalve *Spisula ovalis*. *Oecologia* 111: 331-340.
- Dorenbosch, M., Pollux, B., Pustjens, A., Rajagopal, S., Nagelkerken, I., van der Velde, G., and van der Staay, S. 2006. Population structure of the Dory snapper, *Lutjanus fulvivflamma*, in the western Indian Ocean revealed by means of AFLP fingerprinting. *Hydrobiologia* 568(1): 43-53. doi: 10.1007/s10750-006-0020-8
- Dupanloup, I., Schneider, S., and Excoffier, L. 2002. A simulated annealing approach to define the genetic structure of populations. *Mol Ecol* 11: 2571-2581. doi: 10.1046/j.1365-294X.2002.01650.x
- Excoffier, L., and Lischer, H. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* 10: 564-567. doi: 10.1111/j.1755-0998.2010.02847.x
- Fauvelot, C., Bernardi, G., and Planes, S. 2003. Reductions in the mitochondrial DNA diversity of coral reef fish provide evidence of population bottlenecks resulting from Holocene sea-level change. *Evolution* 57(7): 1571-1583.
- Fisher, R., Leis, J., Clark, D., and Wilson, S. 2005. Critical swimming speeds of late-stage coral reef fish larvae: variation within species, among species and between locations. *Mar Biol* 147(5): 1201-1212. doi: 10.1007/s00227-005-0001-x
- Gaither, M.R., Toonen, R.J., Robertson, D.R., Planes, S., and Bowen, B.W. 2010. Genetic evaluation of marine biogeographical barriers: perspectives from two widespread Indo-Pacific snappers (*Lutjanus kasmira* and *Lutjanus fulvus*). *J Biogeogr* 37(1): 133-147. doi: 10.1111/j.1365-2699.2009.02188.x
- Goudet, J. 1995. FSTAT (vers. 1.2): a computer program to calculate F-statistics. *J Hered* 86: 485-486.
- Grimes, C.B. 1987. Reproductive biology of the Lutjanidae: a review. *In Tropical snappers and groupers: biology and fisheries management. Edited by J.J. Polovina and S. Ralston. Westview Press, Westview, Colorado, USA. pp. 239-294.*
- Hall, T.A. 1999. BioEdit : a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acid Symposium Series* 41: 95-98.
- Hedgecock, D. 1994. Does variance in reproductive success limit effective population size of marine organisms? . *In Genetics and Evolution of Aquatic Organisms. Edited by A.R. Beaumont. Chapman & Hall, London. pp. 122-134.*
- Hogan, J.D., Thiessen, R.J., and Heath, D.D. 2010. Variability in connectivity indicated by chaotic genetic patchiness within and among populations of a marine fish. *Mar Ecol Prog Ser* 417: 263-275. doi: 10.3354/meps08793
- Hudson, R.R. 2000. A New Statistic for Detecting Genetic Differentiation. *Genetics* 155(4): 2011-2014.
- Johnson, M.S., and Black, R. 1982. Chaotic genetic patchiness in an intertidal limpet *Siphonaria* sp. *Mar Biol* 70(2): 157-164. doi: 10.1007/bf00397680
- Jost, L. 2008. G_{ST} and its relatives do not measure differentiation. *Mol Ecol* 17(18): 4015-4026. doi: 10.1111/j.1365-294X.2008.03887.x
- Keeney, D.B., Heupel, M.R., Hueter, R.E., and Heist, E.J. 2005. Microsatellite and mitochondrial DNA analyses of the genetic structure of blacktip shark (*Carcharhinus limbatus*) nurseries in the

northwestern Atlantic, Gulf of Mexico, and Caribbean Sea. *Mol Ecol* 14(7): 1911-1923. doi: 10.1111/j.1365-294X.2005.02549.x

Kelly, R.P., and Palumbi, S.R. 2010. Genetic Structure Among 50 Species of the Northeastern Pacific Rocky Intertidal Community. *PLoS ONE* 5(1): e8594. doi:10.1371/journal.pone.0008594

Larmuseau, M.H.D., Raeymaekers, J.A.M., Hellemans, B., Van Houdt, J.K.J., and Volckaert, F.A.M. 2010. Mito-nuclear discordance in the degree of population differentiation in a marine goby. *Heredity* 105(6): 532-542. doi: 10.1038/hdy.2010.9

Lawton, R.J., Messmer, V., Pratchett, M.S., and Bay, L.K. 2011. High gene flow across large geographic scales reduces extinction risk for a highly specialised coral feeding butterflyfish. *Mol Ecol* 20(17): 3584-3598. doi: 10.1111/j.1365-294X.2011.05207.x

Librado, P., and Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452. doi: 10.1093/bioinformatics/btp187

Molecular Ecology Resources Primer Development Consortium, Bonizzoni, M., Bourjea, J., Chen, B., Crain, B., Cui, L., Fiorentino, V., Hartman, S., Hendricks, S., Ketmaier, V., Ma, X., Muths, D., Pavesi, L., Pfautsch, S., Rieger, M., Santonastaso, T., Sattabongkot, J., Taron, C., Taron, D., Tiedermann, R., Yan, G., Zheng, B., and Zhong, D. 2011. Permanent Genetic Resources added to Molecular Ecology Resources Database 1 April 2011–31 May 2011. *Mol Ecol Res* 11(5): 935-936.

Monsen, K.J., and Blouin, M.S. 2003. Genetic structure in a montane ranid frog: restricted gene flow and nuclear-mitochondrial discordance. *Mol Ecol* 12(12): 3275-3286. doi: 10.1046/j.1365-294X.2003.02001.x

Mora, C., Andrefouet, S., Costello, M.J., Kranenburg, C., Rollo, A., Veron, J., Gaston, K.J., and Myers, R.A. 2006. Ecology. Coral reefs and the global network of Marine Protected Areas. *Science* 312(5781): 1750-1751. doi: 10.1126/science.1125295

Muths, D., Tessier, E., Gouws, G., Craig, M., Mwale, M., Mwaluma, J., Mwandya, A., and Bourjea, J. 2011. Restricted dispersal of the reef fish *Myripristis berndti* at the scale of the South-West Indian Ocean compared to known transoceanic species connectivity. *Mar Ecol Prog Ser* Forthcoming. doi: 10.3354/meps09394

Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nzioka, R.M. 1979. Observations on the spawning seasons of East African reef fishes. *J Fish Biol* 14: 329-342.

Oda, D.K., and Parrish, J. 1987. Trophic ecology of the introduced taape (*Lutjanus kasmira*) in Hawaii. Second International Symposium on Indo-Pacific Marine Biology, Western Society of Naturalists, University of Guam, p. 639.

Palumbi, S.R. 2003. Population genetics, demographic connectivity and the design of marine protected areas. *Ecol Appl* 13: 146-158.

Paul, S. 2000. Efficient genetic markers for population biology. *Trends in Ecology & Evolution* 15(5): 199-203.

Peakall, R., and Smouse, P. 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6: 288-295. doi: 10.1111/j.1471-8286.2005.01155.x

Pothin, K. 2005. Analyse de la dispersion larvaire des poissons récifaux à La Réunion à travers l'étude de leurs otolithes. PhD thesis, Université de La Réunion, Saint-Denis.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155(2): 945-959.

R Development Core Team. 2010. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ragionieri, L., Cannicci, S., Schubart, C., and Fratini, S. 2010. Gene flow and demographic history of the mangrove crab *Neosarmatium meinerti*: A case study from the western Indian Ocean. *Estuar Coast Shelf S* 86: 179-188. doi:10.1016/j.ecss.2009.11.002

Ridgway, T., Hoegh-Guldberg, O., and Ayre, D. 2001. Panmixia in *Pocillopora verrucosa* from South Africa. *Mar Biol* 139(1): 175-181. doi: 10.1007/s002270100573

Ridgway, T., and Sampayo, E.M. 2005. Population Genetic Status of the Western Indian Ocean: What do we Know? *West. Indian Ocean J. Mar. Sci.* 4(1): 1-9.

- Rocha, L.A., Robertson, D.R., Roman, J., and Bowen, B.W. 2005. Ecological speciation in tropical reef fishes. *Proc R Soc Lond B* 272(1563): 573-579. doi: 10.1098/2004.3005
- Rousset, F., and Raymond, M. 1997. Statistical analyses of population genetic data : new tools, old concepts. *Trends Ecol Evol* 12(8): 313-317.
- Saillant, E., Bradfield, S.C., and Gold, J.R. 2010. Genetic variation and spatial autocorrelation among young-of-the-year red snapper (*Lutjanus campechanus*) in the northern Gulf of Mexico. *ICES J. Mar. Sci.* doi: 10.1093/icesjms/fsq011
- Schott, F., Xi, S., and McCreary, P. 2009. Indian Ocean circulation and climate variability. *Rev Geophys* 47: RG1002.
- Schumacher, B., and Parrish, J. 2005. Spatial Relationships Between an Introduced Snapper and Native Goatfishes on Hawaiian Reefs. *Biol. Invasions* 7(6): 925-933. doi: 10.1007/s10530-004-2983-6
- Selkoe, K.A., Gaines, S.D., Caselle, J.E., and Warner, R.R. 2006. Current shifts and kin aggregation explain genetic patchiness in fish recruits. *Ecology* 87(12): 3082-3094.
- Silva, I., Mesquita, N., and Paula, J. 2010. Lack of population structure in the fiddler crab *Uca annulipes* along an East African latitudinal gradient: genetic and morphometric evidence. *Mar Biol* 157: 1113–1126. doi: 10.1007/s00227-010-1393-9
- Song, C.B., Near, T.J., and Page, L.M. 1998. Phylogenetic Relations among Percid Fishes as Inferred from Mitochondrial Cytochrome b DNA Sequence Data. *Mol Phylogenet Evol* 10(3): 343–353.
- Storey, J., and Tibshirani, R. 2003. Statistical significance for genome-wide studies. *Proceedings of the National Academy of Sciences. Proc Natl Acad Sci USA* 100: 9440-9445. doi: 10.1073/pnas.1530509100
- Taberlet, P., Meyer, A., and Bouvet, J. 1992. Unusually large mitochondrial variation in populations of the blue tit, *Parus caeruleus*. *Mol Ecol* 1: 27-36.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3): 585-595.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673-4680.
- Toonen, R.J., Andrews, K., Baums, I., Bird, C., Concepcion, G., Daly-Engel, T., Eble, J., Faucci, A., Gaither, M.R., Iacchei, M., Puritz, J., Schultz, J., Skillings, D., Timmers, M., and Bowen, B.W. 2010. Defining boundaries for applying Ecosystem-based management: A multispecies case study of marine connectivity across the Hawaiian Archipelago. *J Mar Biol.* doi:10.1155/2011/460173
- Van Oosterhout, C., Hutchinson, W., Wills, D., and Shipley, P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes . Mol Ecol Notes* 4: 535-538.
- Visram, S., Yang, M.C., Pillay, R.M., Said, S., Henriksson, O., Grahn, M., and Chen, C.A. 2010. Genetic connectivity and historical demography of the blue barred parrotfish (*Scarus ghobban*) in the western Indian Ocean. *Mar Biol* 157(7): 1475-1487. doi: 10.1007/s00227-010-1422-8
- Weir, B.S., and Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.
- Wright, F. 1969. Volume 2: The theory of gene frequencies. *In Evolution and the genetics of population*. Chicago Press Chicago. p. 512p.

Figures

Figure 1. Geographic location of the sixteen *L. kasmira* localities sampled (with sample size in brackets) and with main oceanic currents indicated by light grey arrows.

Figure 2. Haplotype network representing the evolutionary relationships between mitochondrial haplotypes identified in *L. kasmira*. Connecting lines are proportional to the number of mutational steps between haplotypes (the smallest segment on the figure being for one mutation step). The size of circles is proportional to number of individuals observed for a given haplotype. Colors of pie charts represent geographical origin of the haplotype

Figure 3. Correlograms illustrating the influence of distance on spatial autocorrelation (r) for (3a) mitochondrial data or (3b) microsatellite data for 500km distance class. Confidence Intervals are indicated by dotted lines and confidence error by bars.

Figure 1

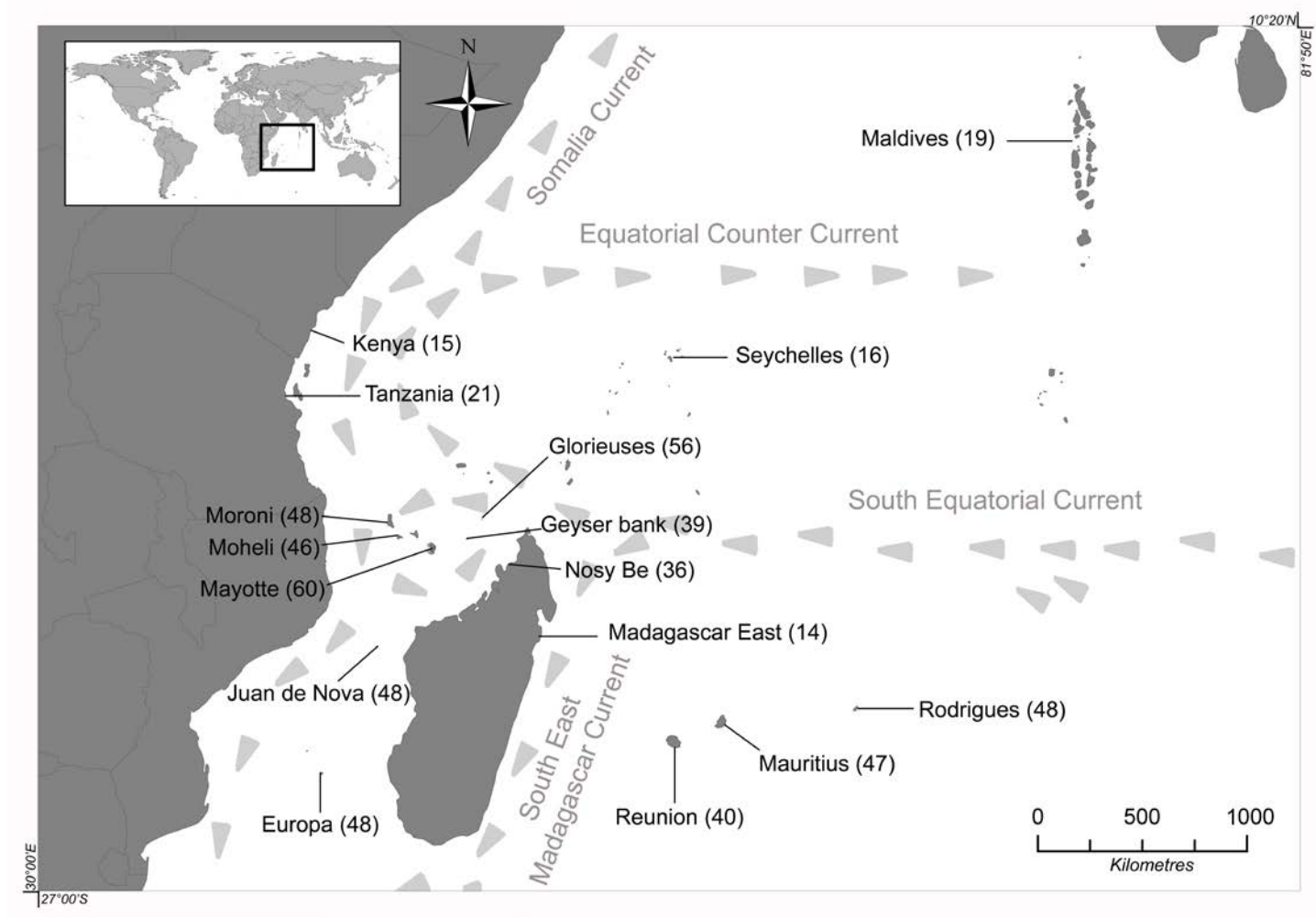


Figure 2

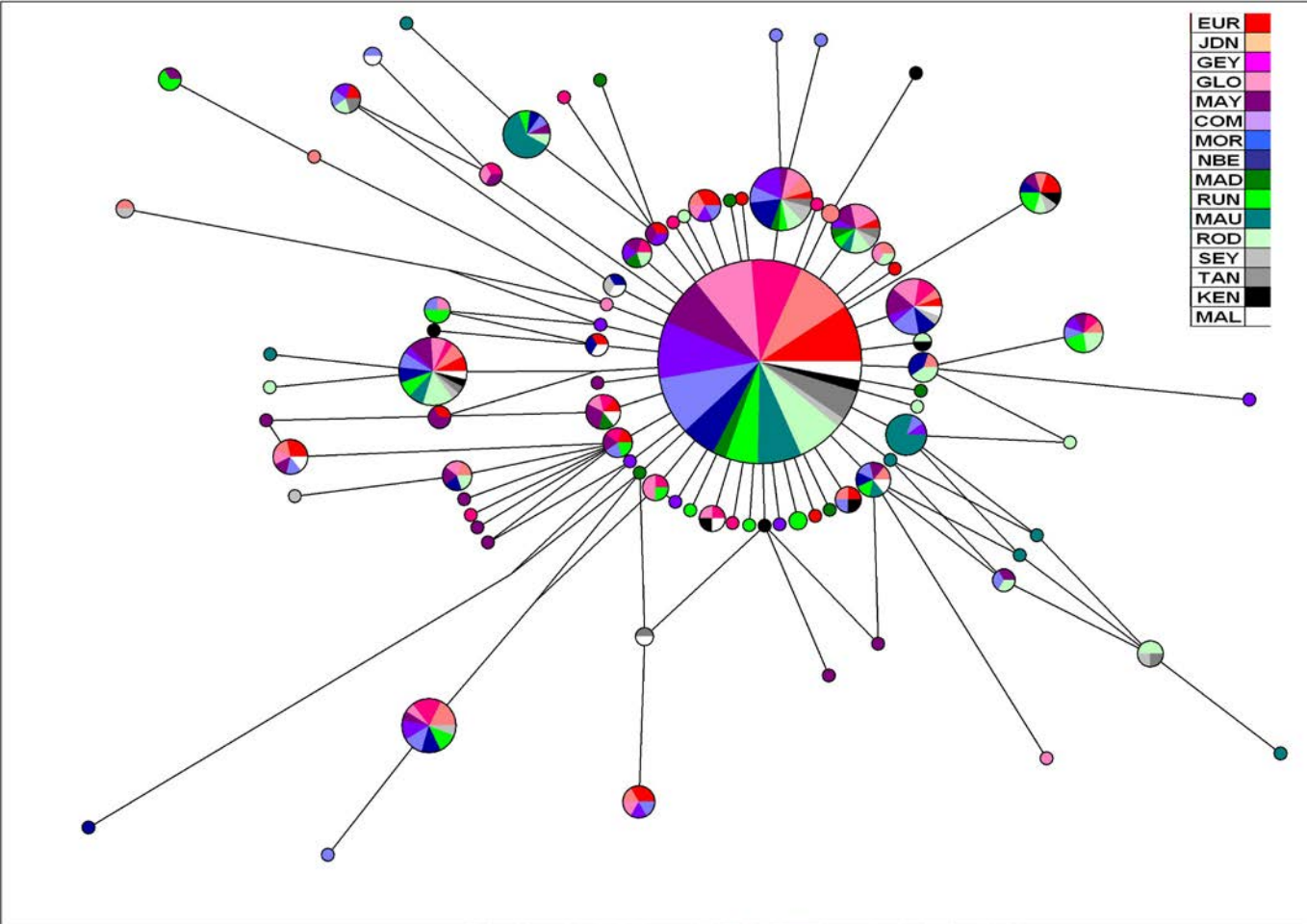
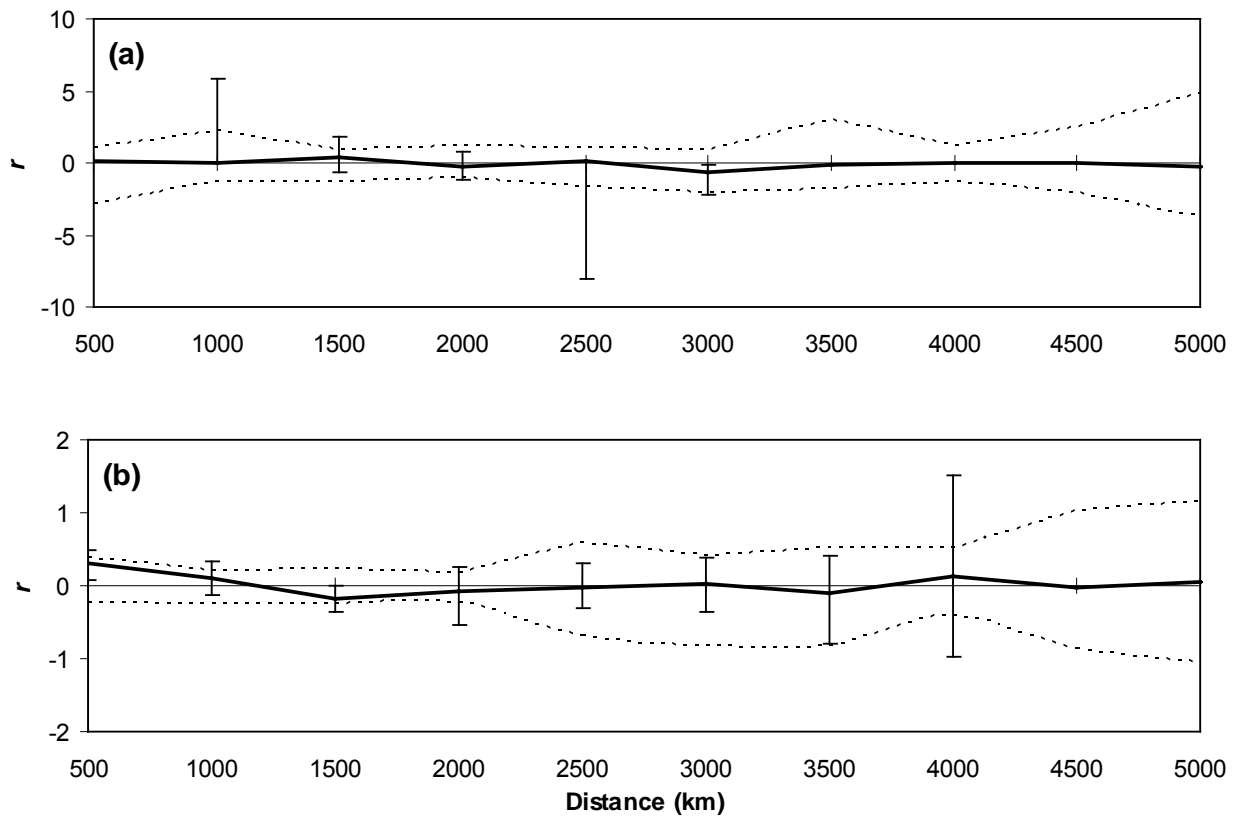


Figure 3



Tables

Table 1. Details of the sixteen localities sampled of *L. kasmira* with sample size for bot mtDNA and microsatellite dataset (N(mtDNA) and N(msat), respectively), GPS coordinates and sampling date. The mean of fish total length have been estimated and indicated in cm.

| Locality | Label | N (mtDNA) | N (msat) | GPS coordinates | | | Sampling date | Mean Fish Total Length (cm \pm SD) | |
|-----------------|-------|--------------|-------------|-----------------|---|-------|---------------|--|------------------|
| Europa | EUR | 46 | 48 | 22°24 | S | 40°23 | E | May-2010 | 23.16 \pm 2.95 |
| Geyser bank | GEY | 39 | 39 | 12°21 | S | 46°26 | E | May-2009 | 20.51 \pm 1.14 |
| Glorieuses | GLO | 50 | 56 | 11°34 | S | 47°23 | E | May-2009 | 18.93 \pm 2.16 |
| Juan de Nova | JDN | 45 | 48 | 17°03 | S | 42°47 | E | May-2010 | 19.65 \pm 1.80 |
| Kenya | KEN | 14 | 15 | 4°65 | S | 39°38 | E | September-2010 | 14.29 \pm 1.53 |
| Madagascar east | MAD | 14 | 14 | 16°50 | S | 49°55 | E | November-2010 | 15.07 \pm 1.59 |
| Maldives | MAL | 19 | 19 | 5°24 | N | 73°15 | E | June-2009 | 20.93 \pm 1.37 |
| Mauritius | MAU | 46 | 47 | 20°16 | S | 57°51 | E | October-2010 | 14.93 \pm 1.99 |
| Mayotte | MAY | 53 | 60 | 12°52 | S | 45°15 | E | March-2008 | 19.7 \pm 1.80 |
| Moheli | MOH | 46 | 46 | 12°24 | S | 43°41 | E | October-2009 | 17.19 \pm 4.44 |
| Moroni | MOR | 48 | 48 | 11°47 | S | 43°14 | E | October-2009 | 15.03 \pm 1.57 |
| Nosy Be | NBE | 35 | 36 | 13°20 | S | 48°15 | E | August-2008 | 19.91 \pm 2.29 |
| Rodrigues | ROD | 47 | 48 | 19°71 | S | 63°42 | E | October-2010 | 15.87 \pm 3.73 |
| Reunion | RUN | 37 | 40 | 21°05 | S | 55°14 | E | March-2007 | 19.89 \pm 5.11 |
| Seychelles | SEY | 16 | 16 | 4°35 | S | 55°27 | E | March-2009 | 18.5 \pm 4.03 |
| Tanzania | TAN | 21 | 21 | 6°48 | S | 39°16 | E | October-2009 | 16.86 \pm 2.56 |

Table 2. Results of AMOVAs made for both markers for the sixteen samples of *L. kasmira* according to different geographical grouping (bold underline significant values):

- (a) 2 groups clustering localities on the West and the East of Madagascar (Group 1= EUR, GEY , GLO , JDN , KEN , MAY , MOH , MOR , NBE , TAN ; Group 2 = MAD, MAU, ROD, RUN, SEY ; Group 0 = MAL)
- (b) 3 groups clustering localities on the North, the West and the East of Madagascar (Group 1 = EUR, GEY , GLO , JDN , MAY , MOH , MOR , NBE ; Group 2 = KEN, MAL, SEY, TAN ; Group 3 = MAD, MAU, ROD, RUN)

| Source of variation | cytochrome <i>b</i> | | | | | Microsatellite | | | | |
|---------------------------------|---------------------|----------------|-------------|----------------|-------------------|----------------|----------------|-------------|----------------|-------------------|
| | d.f. | Sum of squares | % variation | Fixation index | p | d.f. | Sum of squares | % variation | Fixation index | p |
| (a) Among groups | 1 | 3.32 | 1.23 | 0.012 | 0.008 | 1 | 4.04 | 0.04 | 0.000 | 0.27 |
| Among populations within groups | 13 | 12.86 | 0.37 | 0.004 | 0.08 | 13 | 44.76 | 0.71 | 0.007 | < 0.001 |
| Among individuals within pop. | 542 | 471.95 | 98.40 | 0.016 | 0.001 | 1149 | 2564.25 | 99.25 | 0.007 | < 0.001 |
| (b) Among groups | 2 | 4.57 | 1.01 | 0.010 | < 0.001 | 2 | 8.28 | 0.14 | 0.000 | 0.06 |
| Among populations within groups | 13 | 12.26 | 0.22 | 0.002 | 0.26 | 13 | 42.60 | 0.61 | 0.006 | < 0.001 |
| Among individuals within pop. | 560 | 488.90 | 98.77 | 0.012 | < 0.001 | 1184 | 2650.41 | 99.25 | 0.007 | < 0.001 |

Table 3. Cytochrome *b* diversity indices for the sixteen localities of *L. kasmira* and estimated for the whole dataset (with *N* the number of samples analyzed, *Nhap* the number of haplotypes, *h* the haplotype diversity, π the nucleotide diversity, the % of private and main haplotypes, the *D Taj* values (with their associated probability indicated by ** $p < 0.01$ and *** $p < 0.001$), the mismatch distribution parameters τ , θ_0 and θ_1 from which the following demographic parameters were estimated: the age of each population in years and the effective female population size (N_{ef}) was calculated for time zero and present day.

| Locality | <i>N</i> | <i>Nhap</i> | <i>h</i> | π | Private hap. (%) | Main hap. (%) | <i>D Taj.</i> | τ | Population (10^3 years) | age θ_0 | Ne_0 ($*10^3$) | θ_1 | Ne_1 ($*10^3$) |
|--------------------|------------|-------------|--------------|---------------|------------------|---------------|-----------------|-------------|----------------------------|----------------|--------------------|----------------|--|
| EUR | 46 | 18 | 0.706 | 0.0018 | 0.06 | 0.54 | -2.37*** | 1.38 | 95 - 191 | 0.28 | 5 - 10 | 4.34 | 80 - 161 |
| GEY | 39 | 15 | 0.682 | 0.0018 | 0.10 | 0.56 | -2.27*** | 2.14 | 147 - 295 | 0 | 0 | 2.65 | 49 - 98 |
| GLO | 50 | 18 | 0.727 | 0.0019 | 0.04 | 0.52 | -2.45*** | 0.83 | 57 - 114 | 0.77 | 14 - 28 | 13.01 | 241 - 483 |
| JDN | 45 | 16 | 0.690 | 0.0020 | 0.02 | 0.56 | -2.31*** | 3.06 | 210 - 421 | 0 | 0 | 2.43 | 45 - 90 |
| KEN | 14 | 10 | 0.890 | 0.0022 | 0.21 | 0.36 | -2.20*** | 2.26 | 155 - 311 | 0 | 0 | ∞ | ∞ |
| MAD | 14 | 9 | 0.835 | 0.0027 | 0.35 | 0.43 | -2.09** | 1.47 | 101 - 202 | 0 | 0 | ∞ | ∞ |
| MAL | 19 | 12 | 0.836 | 0.0017 | 0 | 0.42 | -2.32** | 2.03 | 139 - 279 | 0.05 | 1 - 2 | 14.63 | 272 - 544 |
| MAU | 46 | 13 | 0.779 | 0.0026 | 0.10 | 0.41 | -1.29 | 2.09 | 143 - 287 | 0 | 0 | 6.78 | 126 - 252 |
| MAY | 53 | 26 | 0.841 | 0.0023 | 0.13 | 0.40 | -2.38*** | 2.33 | 160 - 320 | 0 | 0 | 14.45 | 268 - 537 |
| MOH | 46 | 18 | 0.703 | 0.0028 | 0.10 | 0.54 | -2.47*** | 0.58 | 39 - 79 | 0 | 0 | ∞ | ∞ |
| MOR | 48 | 19 | 0.728 | 0.0024 | 0.06 | 0.52 | -2.39*** | 2.84 | 195 - 390 | 0 | 0 | 3.18 | 59 - 118 |
| NBE | 35 | 14 | 0.806 | 0.0026 | 0.02 | 0.42 | -2.21** | 0.88 | 61 - 122 | 0 | 0 | ∞ | ∞ |
| ROD | 47 | 20 | 0.814 | 0.0025 | 0.08 | 0.42 | -2.11** | 2.13 | 147 - 294 | 0.01 | 0 | 9.81 | 182 - 365 |
| RUN | 37 | 17 | 0.852 | 0.0032 | 0.10 | 0.38 | -2.18*** | 2.38 | 163 - 327 | 0.40 | 7 - 15 | 9.03 | 167 - 335 |
| SEY | 16 | 12 | 0.941 | 0.0042 | 0.06 | 0.25 | -2.25*** | 3.49 | 240 - 480 | 0 | 0 | 24.62 | 457 - 915 |
| TAN | 21 | 8 | 0.623 | 0.0022 | 0 | 0.62 | -1.94** | 3.73 | 256 - 513 | 0 | 0 | 2.11 | 39 - 78 |
| All samples | 576 | 87 | 0.767 | 0.0024 | - | 0.47 | -2.20*** | 2.10 | 144 - 289 | 0.10 | 2 - 4 | 25006.4 | 4.10⁵ - 9.10⁵ |

Table 4. Pairwise mtDNA Φ_{ST} values of differentiation among the sixteen localities of *L. kasmira* (below diagonal) and Jost's D values (above diagonal) Associated p-values are indicated by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Bold indicate values still significant based on Q-values.

| | EUR | GEY | GLO | JDN | KEN | MAD | MAL | MAU | MAY | MOH | MOR | NBE | ROD | RUN | SEY | TAN |
|-----|---------|--------|-----------------|-----------------|--------|--------|--------|---------|--------|--------|--------|--------|--------|--------|-------|-------|
| EUR | -- | 0 | 0 | 0 | 0.015 | 0 | 0 | 0.114 | 0.010 | 0 | 0 | 0.012 | 0.002 | 0.042 | 0.178 | 0 |
| GEY | -0.006 | -- | 0 | 0 | 0.054 | 0.001 | 0 | 0.130 | 0.026 | 0 | 0 | 0.017 | 0.030 | 0.049 | 0.208 | 0 |
| GLO | -0.009 | -0.009 | -- | 0 | 0.015 | 0 | 0 | 0.113 | 0 | 0 | 0 | 0 | 0 | 0.025 | 0.121 | 0 |
| JDN | -0.009 | -0.012 | -0.007 | -- | 0.030 | 0 | 0 | 0.125 | 0.028 | 0 | 0 | 0 | 0.011 | 0.035 | 0.164 | 0 |
| KEN | 0.006 | 0.017 | 0.005 | 0.011 | -- | 0 | 0 | 0.105 | 0 | 0.045 | 0.017 | 0 | 0 | 0 | 0 | 0.090 |
| MAD | -0.006 | 0.001 | -0.013 | -0.000 | -0.018 | -- | 0 | 0.086 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.006 |
| MAL | -0.008 | -0.002 | -0.011 | -0.003 | -0.029 | -0.024 | -- | 0.084 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.024 |
| MAU | 0.038** | 0.045* | 0.035*** | 0.043*** | 0.021 | 0.019 | 0.019 | -- | 0.101 | 0.110 | 0.087 | 0.103 | 0.092 | 0.106 | 0.233 | 0.127 |
| MAY | 0.002 | 0.007 | -0.001 | 0.008 | -0.013 | -0.017 | -0.018 | 0.022* | -- | 0.025 | 0 | 0 | 0 | 0 | 0 | 0.066 |
| MOH | -0.008 | -0.008 | -0.009 | -0.011 | 0.014 | -0.009 | 0.000 | 0.037** | 0.007 | -- | 0 | 0 | 0.006 | 0.040 | 0.147 | 0 |
| MOR | -0.009 | -0.011 | -0.010 | -0.012 | 0.005 | -0.007 | -0.011 | 0.027 | -0.001 | -0.011 | -- | 0 | 0 | 0.016 | 0.136 | 0 |
| NBE | 0.004 | 0.005 | -0.002 | -0.002 | -0.007 | -0.015 | -0.015 | 0.026 | -0.006 | -0.004 | -0.006 | -- | 0 | 0 | 0 | 0.033 |
| ROD | 0.000 | 0.009 | -0.000 | 0.003 | -0.013 | -0.015 | -0.010 | 0.023 | -0.007 | 0.001 | -0.001 | -0.008 | -- | 0 | 0.007 | 0.026 |
| RUN | 0.012 | 0.014 | 0.007 | 0.010 | -0.015 | -0.012 | -0.009 | 0.023 | -0.007 | 0.011 | 0.004 | -0.005 | -0.006 | -- | 0 | 0.088 |
| SEY | 0.043* | 0.052* | 0.028 | 0.042 | -0.015 | -0.010 | -0.004 | 0.040* | -0.002 | 0.036 | 0.031 | -0.011 | 0.001 | -0.006 | -- | 0.202 |
| TAN | -0.011 | -0.006 | -0.010 | -0.009 | 0.029 | 0.001 | 0.006 | 0.046* | 0.017 | -0.012 | -0.007 | 0.009 | 0.005 | 0.025 | 0.055 | -- |

Table 5. Microsatellite diversity indices for the sixteen localities of *L. kasmira* (with N the number of samples analyzed, Nall the average number of alleles per population, Rs the allelic richness, Ho and He, respectively, the observed and expected heterozygosities and Fis the fixation index with associated significance * p < 0.05, *** p < 0.001 Details are given per loci in Supplementary materials)

| Locality | N | Nall | Rs | Ho | He | Fis |
|-----------------|------------|-------------|-------------|--------------|--------------|-----------------|
| EUR | 48 | 8.0 | 6.19 | 0.544 | 0.602 | 0.098*** |
| GEY | 39 | 7.8 | 6.06 | 0.480 | 0.553 | 0.134*** |
| GLO | 56 | 7.6 | 5.72 | 0.450 | 0.536 | 0.161*** |
| JDN | 48 | 7.9 | 5.98 | 0.440 | 0.560 | 0.217*** |
| KEN | 15 | 6.5 | 6.59 | 0.516 | 0.650 | 0.212*** |
| MAD | 14 | 6.5 | 6.62 | 0.517 | 0.576 | 0.105 |
| MAL | 19 | 6.7 | 6.51 | 0.534 | 0.615 | 0.135* |
| MAU | 47 | 7.0 | 5.45 | 0.417 | 0.507 | 0.179*** |
| MAY | 60 | 7.7 | 5.68 | 0.470 | 0.532 | 0.117*** |
| MOH | 46 | 7.9 | 5.92 | 0.472 | 0.551 | 0.144*** |
| MOR | 48 | 7.9 | 6.16 | 0.458 | 0.537 | 0.148*** |
| NBE | 36 | 7.2 | 6.02 | 0.534 | 0.598 | 0.108*** |
| ROD | 48 | 7.7 | 5.69 | 0.489 | 0.541 | 0.096*** |
| RUN | 40 | 8.0 | 6.52 | 0.503 | 0.608 | 0.175*** |
| SEY | 16 | 6.5 | 6.21 | 0.546 | 0.568 | 0.039 |
| TAN | 21 | 6.3 | 5.72 | 0.476 | 0.560 | 0.154*** |
| TOTAL | 601 | 7.32 | 6.17 | 0.441 | 0.595 | 0.141*** |

Table 6. Pairwise values of differentiation between the sixteen localities of *L. kasmira*. Microsatellite F_{ST} values are below the diagonal and Jost's D values are above the diagonal. Associated p-values are indicated by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

| | EUR | GEY | GLO | JDN | KEN | MAD | MAL | MAU | MAY | MOH | MOR | NBE | ROD | RUN | SEY | TAN |
|-----|----------|---------|----------|----------|---------|-------|--------|----------|----------|-------|----------|---------|--------|-------|-------|-------|
| EUR | -- | 0 | 0 | 0.002 | 0.003 | 0.004 | 0.006 | 0 | 0 | 0 | 0.002 | 0 | 0 | 0.003 | 0 | 0.004 |
| GEY | 0.002 | -- | 0 | 0.005 | 0.001 | 0.001 | 0 | 0 | 0.009 | 0.007 | 0.003 | 0.002 | 0.001 | 0.001 | 0 | 0.006 |
| GLO | 0.013*** | 0.001 | -- | 0 | 0.002 | 0.004 | 0.002 | 0 | 0 | 0 | 0 | 0 | 0 | 0.001 | 0 | 0 |
| JDN | 0.004 | 0.001 | 0.004 | -- | 0.003 | 0 | 0.001 | 0 | 0.009 | 0.004 | 0.005 | 0.009 | 0 | 0.007 | 0 | 0.004 |
| KEN | 0.007 | 0.005 | 0.010 | 0.014 | -- | 0.016 | 0.006 | 0.005 | 0.007 | 0.005 | 0.013 | 0.005 | 0.005 | 0.003 | 0.005 | 0.007 |
| MAD | 0.001 | 0.002 | 0.006 | 0.008 | 0.005 | -- | 0.001 | 0 | 0.001 | 0 | 0.016 | 0.014 | 0.001 | 0.003 | 0.003 | 0 |
| MAL | 0 | 0 | 0.004 | 0.005 | 0 | 0 | -- | 0 | 0.011 | 0.002 | 0.005 | 0.010 | 0.003 | 0.009 | 0.005 | 0.012 |
| MAU | 0.010*** | 0.002 | 0.011** | 0.009* | 0.018* | 0.012 | 0.008 | -- | 0.003 | 0.001 | 0 | 0 | 0 | 0 | 0 | 0 |
| MAY | 0.011*** | 0 | 0.003 | 0.006 | 0.008 | 0.004 | 0.004 | 0.006 | -- | 0.002 | 0.019 | 0.011 | 0.003 | 0.009 | 0.001 | 0.006 |
| MOH | 0.004 | 0 | 0.001 | 0.003 | 0.008 | 0.005 | 0 | 0.000 | 0.000 | -- | 0.008 | 0.004 | 0.004 | 0.010 | 0.007 | 0.016 |
| MOR | 0.021*** | 0.009** | 0.006** | 0.017*** | 0.021** | 0.008 | 0.012* | 0.019*** | 0.010*** | 0.006 | -- | 0.000 | 0.004 | 0.009 | 0 | 0.005 |
| NBE | 0.003 | 0.003 | 0.010** | 0.004 | 0.011 | 0.006 | 0 | 0.013* | 0.007* | 0.002 | 0.012*** | -- | 0.001 | 0.002 | 0.001 | 0.011 |
| ROD | 0.002 | 0 | 0.004 | 0.006 | 0.009 | 0.004 | 0.003 | 0.002 | 0.005 | 0 | 0.018*** | 0.007** | -- | 0.008 | 0.002 | 0.005 |
| RUN | 0.002 | 0.003 | 0.011*** | 0.006 | 0.004 | 0.008 | 0 | 0.013*** | 0.011*** | 0.006 | 0.021*** | 0.003 | 0.006* | -- | 0 | 0.002 |
| SEY | 0.004 | 0.007 | 0.009 | 0.012 | 0.007 | 0.005 | 0 | 0.008 | 0.010 | 0.003 | 0.010 | 0.005 | 0.009 | 0.001 | -- | 0.004 |
| TAN | 0.009 | 0.005 | 0.007 | 0.012 | 0.004 | 0.010 | 0.002 | 0.010 | 0.013* | 0.004 | 0.005 | 0.007 | 0.009 | 0.011 | 0.006 | -- |

Supplementary materials: Table A. Microsatellite information detailed per locus for the sixteen localities of *L. kasmira* (with N all the number of alleles per population, Ho and He, respectively, the observed and expected heterozygosities and associated significance * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns for non significant)

| | | Locus | | | | | | | |
|---------------|----|-------|-------|-------|-------|-------|-------|-------|-------|
| | | Lk-08 | Lk-10 | Lk-18 | Lk-27 | Lk-29 | Lk-30 | Lk-49 | Lk-57 |
| EUR N = 48 | Na | 3 | 10 | 5 | 16 | 6 | 14 | 2 | 8 |
| | Ho | 0,188 | 0,771 | 0,271 | 0,875 | 0,625 | 0,792 | 0,292 | 0,542 |
| | He | 0,243 | 0,813 | 0,384 | 0,913 | 0,667 | 0,903 | 0,305 | 0,543 |
| | p | ns | ns | ns | ns | *** | ** | ns | ns |
| GEY N = 39 | Na | 3 | 8 | 4 | 15 | 5 | 15 | 4 | 7 |
| | Ho | 0,256 | 0,769 | 0,231 | 0,744 | 0,333 | 0,769 | 0,231 | 0,513 |
| | He | 0,229 | 0,811 | 0,231 | 0,920 | 0,512 | 0,894 | 0,249 | 0,530 |
| | p | *** | * | ns | ns | ns | *** | ns | *** |
| GLO N = 56 | Na | 3 | 8 | 3 | 15 | 4 | 15 | 3 | 7 |
| | Ho | 0,125 | 0,643 | 0,161 | 0,821 | 0,375 | 0,804 | 0,304 | 0,375 |
| | He | 0,287 | 0,813 | 0,150 | 0,908 | 0,373 | 0,901 | 0,287 | 0,534 |
| | p | ns | ns | ns | ns | *** | ** | ns | *** |
| JDN N = 48 | Na | 3 | 8 | 4 | 16 | 7 | 13 | 2 | 10 |
| | Ho | 0,271 | 0,708 | 0,250 | 0,750 | 0,354 | 0,604 | 0,229 | 0,354 |
| | He | 0,321 | 0,811 | 0,241 | 0,906 | 0,563 | 0,893 | 0,234 | 0,469 |
| | p | *** | ns | *** | ns | ns | ns | ** | *** |
| KEN N = 15 | Na | 4 | 6 | 5 | 14 | 4 | 11 | 4 | 6 |
| | Ho | 0,133 | 0,800 | 0,333 | 0,867 | 0,467 | 0,733 | 0,333 | 0,467 |
| | He | 0,436 | 0,796 | 0,438 | 0,898 | 0,509 | 0,880 | 0,429 | 0,647 |
| | p | ** | ns | ns | ns | ns | ns | ns | ** |
| MAD N = 14 | Na | 3 | 8 | 2 | 12 | 7 | 12 | 3 | 6 |
| | Ho | 0,071 | 0,857 | 0,286 | 0,857 | 0,571 | 0,786 | 0,357 | 0,357 |
| | He | 0,196 | 0,832 | 0,245 | 0,867 | 0,597 | 0,855 | 0,390 | 0,464 |
| | p | *** | ns | ns | ns | *** | ns | ns | ns |
| MAL N = 19 | Na | 3 | 7 | 4 | 14 | 5 | 14 | 3 | 6 |
| | Ho | 0,167 | 0,944 | 0,333 | 0,833 | 0,278 | 0,889 | 0,278 | 0,556 |
| | He | 0,329 | 0,779 | 0,292 | 0,906 | 0,576 | 0,910 | 0,387 | 0,608 |
| | p | *** | ** | *** | *** | ns | ns | ns | ns |
| MAU N = 47 | Na | 2 | 8 | 3 | 17 | 4 | 14 | 3 | 5 |
| | Ho | 0,000 | 0,596 | 0,128 | 0,723 | 0,532 | 0,766 | 0,213 | 0,383 |
| | He | 0,042 | 0,800 | 0,231 | 0,921 | 0,517 | 0,884 | 0,256 | 0,369 |
| | p | ns | ns | ns | ns | *** | ns | *** | ns |
| MAY N = 60 | Na | 3 | 9 | 4 | 14 | 7 | 16 | 4 | 6 |
| | Ho | 0,250 | 0,733 | 0,133 | 0,800 | 0,283 | 0,814 | 0,267 | 0,483 |
| | He | 0,245 | 0,807 | 0,156 | 0,906 | 0,459 | 0,899 | 0,285 | 0,466 |
| | p | ns | ns | ns | ns | *** | * | ** | *** |
| MOH N = 46 | Na | 3 | 8 | 5 | 18 | 4 | 14 | 3 | 7 |
| | Ho | 0,196 | 0,609 | 0,152 | 0,848 | 0,413 | 0,804 | 0,326 | 0,435 |
| | He | 0,213 | 0,801 | 0,234 | 0,922 | 0,472 | 0,890 | 0,305 | 0,527 |
| | p | * | ns | ns | ns | ns | ** | ns | ** |
| MOR N = 48 | Na | 2 | 8 | 4 | 16 | 5 | 16 | 5 | 9 |
| | Ho | 0,146 | 0,625 | 0,229 | 0,938 | 0,125 | 0,833 | 0,438 | 0,333 |
| | He | 0,170 | 0,778 | 0,276 | 0,919 | 0,229 | 0,910 | 0,475 | 0,496 |
| | p | ns | ns | ns | ns | ns | ** | *** | ns |
| NBE N = 36 | Na | 4 | 8 | 3 | 14 | 7 | 13 | 4 | 6 |
| | Ho | 0,333 | 0,722 | 0,306 | 0,806 | 0,444 | 0,806 | 0,417 | 0,444 |
| | He | 0,345 | 0,783 | 0,263 | 0,910 | 0,628 | 0,888 | 0,459 | 0,446 |
| | p | ns | ns | ns | ns | *** | *** | *** | ns |
| ROD N = 48 | Na | 3 | 8 | 3 | 15 | 6 | 15 | 4 | 6 |
| | Ho | 0,167 | 0,667 | 0,250 | 0,854 | 0,500 | 0,833 | 0,167 | 0,479 |

| | | | | | | | | | |
|------------|-----------|-------|-------|-------|-------|-------|-------|-------|-------|
| | He | 0,154 | 0,797 | 0,221 | 0,902 | 0,595 | 0,902 | 0,225 | 0,488 |
| | p | ns | ns | ns | ns | *** | *** | *** | ns |
| RUN | Na | 6 | 9 | 3 | 15 | 6 | 16 | 2 | 8 |
| N = 40 | Ho | 0,075 | 0,750 | 0,175 | 0,800 | 0,625 | 0,775 | 0,300 | 0,525 |
| | He | 0,370 | 0,820 | 0,243 | 0,908 | 0,657 | 0,919 | 0,320 | 0,571 |
| | p | *** | ns | ns | ns | ns | ns | ns | ns |
| SEY | Na | 2 | 6 | 2 | 14 | 5 | 14 | 4 | 5 |
| N = 16 | Ho | 0,125 | 0,750 | 0,125 | 0,813 | 0,563 | 0,938 | 0,563 | 0,500 |
| | He | 0,117 | 0,736 | 0,117 | 0,908 | 0,594 | 0,900 | 0,463 | 0,570 |
| | p | ns | ns | ns | ns | ns | ns | ns | * |
| TAN | Na | 2 | 7 | 4 | 14 | 4 | 11 | 4 | 4 |
| N = 21 | Ho | 0,143 | 0,762 | 0,286 | 0,762 | 0,333 | 0,762 | 0,333 | 0,429 |
| | He | 0,133 | 0,802 | 0,290 | 0,910 | 0,423 | 0,870 | 0,523 | 0,427 |
| | p | ns | ** | ns | ns | ns | ns | ns | ns |