

ORIGINAL ARTICLE

Speciation with gene flow in the large white-headed gulls: does selection counterbalance introgression?

L Gay¹, G Neubauer², M Zagalska-Neubauer², J-M Pons³, DA Bell⁴ and P-A Crochet¹¹CEFE-CNRS, UMR 5175, Montpellier cedex 5, France; ²Ornithological Station, Museum and Institute of Zoology, Polish Academy of Sciences, Gdańsk, Poland; ³Origine, Structure et Evolution de la biodiversité, UMR 5202, Département Systématique et Evolution, Muséum National d'Histoire Naturelle, Paris, France and ⁴Department of Ornithology and Mammalogy, California Academy of Sciences, San Francisco, CA, USA

We investigated the role of selection in generating and maintaining species distinctness in spite of ongoing gene flow, using two zones of secondary contact between large gull species in Europe (*Larus argentatus* and *Larus cachinnans*) and North America (*Larus glaucescens* and *Larus occidentalis*). We used the pattern of neutral genetic differentiation at nine microsatellite loci (F_{ST}) as an indicator of expected changes under neutral processes and compared it with phenotypic differentiation (P_{ST}) for a large number of traits (size, plumage melanism and coloration of bare parts). Even assuming very low heritability, interspecific divergence between *L. glaucescens* and *L. occidentalis* in plumage melanism and orbital ring colour clearly exceeded neutral differentiation. Similarly, melanism of the central primaries was highly divergent between *L. argentatus* and *L. cachin-*

nans. Such divergence is unlikely to have arisen randomly and is therefore attributed to spatially varying selection. Variation in plumage melanism in both transects agrees with Gloger's rule, which suggests that latitude (and associated sun and humidity gradients) could be the selective pressure shaping differentiation in plumage melanism. We suggest that strong species differentiation in orbital ring colour results from sexual selection. We conclude that these large gull species, along with other recently diverged species that hybridize after coming into secondary contact, may differ only in restricted regions of the genome that are undergoing strong disruptive selection because of their phenotypic effects.

Heredity (2009) **102**, 133–146; doi:10.1038/hdy.2008.99; published online 24 September 2008

Keywords: selection; drift; F_{ST} ; P_{ST} ; Q_{ST} ; microsatellites; phenotypic divergence; *Larus*

Introduction

Differential adaptation to natural or sexual environments can drive phenotypic divergence of populations, sometimes without any geographic isolation (Via, 2001, 2002; Kirkpatrick and Ravigne, 2002; Gavrillets, 2003). Although central to the models of sympatric speciation (Johnson and Gullberg, 1998), this concept is also essential for determining the effect on phenotypic differentiation when allopatric populations come into secondary contact. However, the extent to which natural selection and the adaptive responses of organisms can override the effects of gene flow and maintain species distinctness in spite of ongoing gene flow is far from clarified (Ehrlich and Raven, 1969; Schluter, 2001; Doebeli *et al.*, 2005).

In the absence of selection, genetic drift between allopatric populations drives divergence of the whole nuclear genome at the same rate. When gene flow is not completely interrupted (incomplete reproductive isolation), it affects all genes to the same degree. Natural

selection, on the other hand, increases the rate of divergence for loci under selection and acts against gene flow at those loci more than at neutral loci. As a result, the genes under selection (for example, those involved in local adaptation) may show a greater level of differentiation than the rest of the genome (Wu, 2001). A F_{ST} -based survey of genetically mapped loci might thus highlight the different genomic regions that are subject to strong disruptive selection or assortative mating (Beaumont, 2005) and help quantify levels of adaptive differentiation among populations and recently derived species. For example, in *Drosophila pseudoobscura* and some close relatives, some nuclear molecular markers vary greatly in their level of differentiation (Wang *et al.*, 1997), an observation that is not compatible with a model of simple allopatric speciation. This species complex thus appears to fit a speciation model in which species continue to exchange genes at some loci and not at others (Machado *et al.*, 2002; Noor *et al.*, 2007). Comparing levels of differentiation between neutral markers and quantitative traits thus provides insight into the role of selection in maintaining phenotypic variation and ultimately into speciation mechanisms (Lande, 1992; Merilä and Crnokrak, 2001; Reed and Frankham, 2001; McKay and Latta, 2002). This approach can be used to test a null hypothesis of no influence of selection by determining whether the level of trait divergence attributable to additive genetic effects (as measured by

Correspondence: Dr L Gay, Centre for Ecology and Conservation, University of Exeter, Cornwall Campus, Tremough, Penryn TR10 9EZ, UK.

E-mail: L.Gay@exeter.ac.uk

Received 4 February 2008; revised 28 July 2008; accepted 15 August 2008; published online 24 September 2008

Q_{ST}) equals the level of divergence at neutral loci (as measured by F_{ST}) (Spitze, 1993; Steinger *et al.*, 2002; Edmands and Harrison, 2003; Sanou *et al.*, 2005; Volis *et al.*, 2005). Although trait divergence should be measured as the among-population proportion of additive genetic variance (Q_{ST} , Leinonen *et al.*, 2008), the total phenotypic variance is often used as a surrogate when environmental and non-additive genetic effects cannot be separated. This index of phenotypic divergence is called P_{ST} (for 'phenotypic Q_{ST} ' or 'pseudo- Q_{ST} ', Saether *et al.*, 2007). P_{ST} is particularly useful for organisms for which a quantitative genetics design is not possible, as long as the underlying assumptions are understood and discussed (Merilä *et al.*, 1997; Storz, 2002; Saint-Laurent *et al.*, 2003; Raeymaekers *et al.*, 2007).

The large white-headed gulls constitute a promising model to study whether selection can maintain species distinctness in spite of ongoing gene flow. Although interspecific genetic differentiation at neutral nuclear loci is surprisingly low (microsatellites F_{ST} between 5–10%; Crochet *et al.*, 2003), several phenotypic characters exhibit clear-cut differences even between those species that are most similar in allozyme or microsatellite allele frequency. Genetic differentiation is also higher in the mitochondrial than in the nuclear genome, which has been explained by a combination of recent species origin and ongoing interspecific gene flow (Crochet *et al.*, 2003). Indeed, interspecific hybrids between large gull species have frequently been recorded (Pierotti, 1987; Olsen and Larsson, 2004; Gay *et al.*, 2007). There is thus imperfect reproductive isolation between large white-headed gull species, with ongoing interspecific gene flow counteracting the effect of genetic drift. The very different situation for phenotypic traits suggests that selection is important in generating and maintaining species-specific differences in this group.

The aim of this study was to determine whether evidences of selection could explain phenotypic differences between large gull species. We investigated differentiation in phenotypic traits and neutral markers within and between species in two zones of secondary contact (two replicates differing in their respective level of divergence) between *Larus argentatus* and *Larus cachinnans* in Europe and between *Larus glaucescens* and *Larus occidentalis* in North America. We used patterns of neutral genetic differentiation at nine microsatellite loci (F_{ST}) as an indicator of expected changes under neutral processes and compared it with phenotypic differentiation (P_{ST}) for a large number of traits (size, plumage melanism and coloration of bare parts). We expect evidence of selection only on a subset of traits because a large number of loci under selection in the genome would counteract the homogenizing action of gene flow on all neutral markers through genetic linkage.

Materials and methods

Study sites: two secondary contacts between large gulls *L. argentatus* and *L. cachinnans* came into secondary contact in Western Europe at the end of the 20th century, following concomitant range expansion southwards (Neubauer *et al.*, 2006) and westwards, respectively (Faber *et al.*, 2001). Field observations (Panov and Monzиков, 1999), combined with a recent study based

on molecular markers (Gay *et al.*, 2007), demonstrated that hybridization and introgression occur across this secondary contact zone. Four populations were sampled along a transect spanning the range of both *L. argentatus* and *L. cachinnans*, including their zone of secondary contact (Figure 1a). The population sampled in northern Poland (Gdynia) represents pure *L. argentatus* only, whereas populations sampled in southern Poland (Tarnow) and Ukraine (Molochnyy Lyman) belong to *L. cachinnans* only. Both species occur in the population from central Poland (Wloclawek). Samples were taken from adult gulls trapped on the nest during the breeding season from 1998 to 2004 and included blood stored in EDTA buffer, growing feather quills or muscle tissue preserved in ethanol (Table 1). These data are a subset of the data used by Gay *et al.* (2007) to estimate molecular and morphological patterns of introgression in this zone of secondary contact. Only the populations for which morphological data were available were included (four out of seven).

The zone of secondary contact between *L. glaucescens* and *L. occidentalis* lies along coastal Oregon and Washington, at the southern range limit of *L. glaucescens* and the northern range limit of *L. occidentalis*. Hybridization has been reported between these taxa since the beginning of the twentieth century. Previous studies suggested weak or no apparent barriers to gene flow between the two taxa within the 180-km long hybridization area (Bell, 1996, 1997; Good *et al.*, 2000). Samples and morphological measurements were collected from 1985 to 1990 during the breeding season in coastal Alaska, British Columbia, Washington, Oregon and California (see Bell, 1996 for sampling details). All adult specimens were collected on nests or from the sea or land points near colonies. They were measured for qualitative and quantitative morphological traits in all 12 populations (Figure 1b). Tissue samples, including liver, heart, kidney or pectoral muscle, were frozen in liquid nitrogen and stored at -70° . Tissue samples for DNA analysis were later transferred to ethanol (Table 1). All individuals were sexed by looking at the gonads and measured by the same observer (DAB).

Morphological measurements

In Europe, 20 quantitative or ordinal variables characterizing morphometry, plumage melanism and coloration of bare parts were measured on adults in the four populations. Seven morphometric variables were measured with a calliper to the nearest millimetre, namely tarsus, toe, wing, head and bill length, bill depth at gonys and minimum bill depth. We characterized plumage melanism using the pattern of the outer and the central primaries. The outer primaries were described by four quantitative variables (length of black and of white on the 9th and 10th primary) and one ordinal variable (colour of the tongue on the inner web of the 10th primary). The central primaries were described by three quantitative variables (number of primaries with black, length of black on the 7th and 8th primary). The coloration of bare parts was characterized by five ordinal variables describing the leg colour (tarsus, toe and web colour), the orbital ring colour and the amount of dark spots in the iris. More details about morphological measurements can be found in Gay *et al.* (2007).

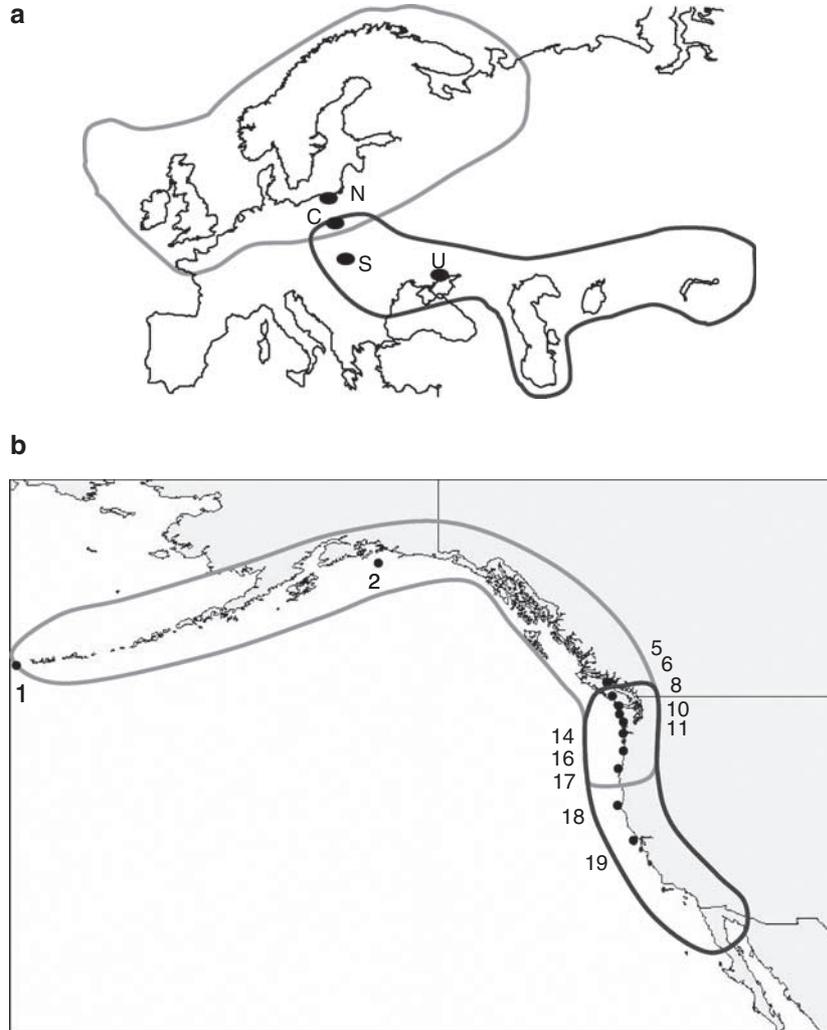


Figure 1 Distribution range and sampling sites for two secondary contact zones between large gulls (Olsen and Larsson, 2004): (a) in Europe between herring gull (*Larus argentatus*, in light grey) and Caspian gull (*Larus cachinnans*, in dark grey); N: northern Poland (Gdynia); C: central Poland (Włocławek); S: southern Poland (Tarnow); U: Ukraine (Molochnyy Lyman). (b) In North America between glaucous-winged gull (*Larus glaucescens*, in light grey) and western gull (*Larus occidentalis*, in dark grey). Sampled populations were numbered from 1 to 19; 1: Aleutian Island (AL, USA); 2: Kachemak Bay (AL, USA); 5: Vancouver Island North (BC, Canada); 6: Vancouver Island South (BC, Canada); 8: Tatoosh Island (WA, USA); 10: Destruction Island (WA, USA); 11: Grays Harbor, Goose Island (WA, USA); 14: Baker Bay at Columbia River mouth Clatsop, East Sand Island (OR, USA); 16: Yaquina Bay Lincoln (OR, USA); 17: Coos Bay near Coos head (OR, USA); 18: Sugarloaf Rock at Cape Mendocino Humboldt (CA, USA); 19: San Francisco Bay, Alcatraz Island (CA, USA).

Repeatability was significant and high for all traits, from 75 to 100% (Lessells and Boag, 1987). We summarized these highly correlated variables using multivariate analyses (PCA) on size, plumage melanism (separated into outer and central primaries) and leg colour in ADE-4 (Thioulouse *et al.*, 1997) (Table 2). We used orbital ring colour and the amount of dark spots in the iris without transformation.

In North America, 14 morphometric variables were measured with a calliper to the nearest millimetre, namely length of wing, tail, 8th, 9th and 10th primary feathers, tarsus, middle toe, head, culmen and bill, bill width, bill depth at gonys, bill depth at nares (anterior border) and bill depth at nares (posterior border). To characterize the plumage melanism, we measured the coloration of the mantle and the dark elements of primary tips on wing preparations, and studied skins with a Munsell 37-step neutral value scale (Munsell, 1971). Five traits were measured to describe the coloration

of bare parts, namely coloration of bill, iris, dark iris spots and orbital ring (measured independently for each eye). Measurements were made on freshly collected specimens using hand-held Munsell colour charts (matte finish) from the *Munsell Book of Colour* (Munsell, 1976). Each Munsell colour is defined by three characters—hue, value and chroma—from which dominant wavelength, percentage of spectral reflectance or brightness and excitation purity can be obtained with the aid of conversion tables (Munsell, 1968) (see Bell, 1996 for additional details on morphological measurements). We thus obtained a total of 15 measurements for the coloration of bare parts for each bird. As for the European data set, we used PCAs to summarize variables on size, shape, plumage melanism, orbital ring, bill, iris and dark iris spots colour. Table 2 recapitulates the correlated variables we grouped and summarized by PCA to form synthetic variables.

Table 1 Summary statistics of molecular markers and phenotypic variables studied in 4 populations across the zone of secondary contact between *Larus argentatus* and *Larus cachinnans* and 12 populations across the zone of secondary contact between *Larus glaucescens* and *Larus occidentalis*

	Population	Phenotype	Microsatellites			
			n_{morpho}	n_{micro}	$n_{alleles}$	H_e
Europe	Northern Poland	28	20	4.7	0.50 (0.29)	0.11*
	Central Poland	138	106	6.9	0.62 (0.24)	0.11**
	Southern Poland	80	20	5.2	0.55 (0.28)	0.10*
	Ukraine	57	32	5.4	0.56 (0.27)	0.11**
North America	1	10	10	3.2	0.42 (0.30)	-0.02
	2	31	28	5.6	0.57 (0.20)	0.12
	5	21	23	5.1	0.56 (0.27)	0.004
	6	26	26	5.4	0.62 (0.21)	0.25**
	8	40	35	5.7	0.54 (0.28)	-0.01
	10	18	18	4.9	0.58 (0.21)	0.11**
	11	35	18	5.0	0.56 (0.20)	-0.06
	14	23	27	5.2	0.51 (0.25)	0.03
	16	18	17	4.3	0.52 (0.27)	0.13
	17	27	27	4.1	0.51 (0.29)	0.07
	18	53	18	3.9	0.48 (0.25)	0.05
	19	37	26	3.8	0.46 (0.28)	0.09

Sample size for morphological measurements (n_{morpho}), sample size (n_{micro}), average number of alleles per population ($n_{alleles}$), expected heterozygosity (H_e with standard error in parentheses) and global F_{IS} for the nine microsatellite loci are given in the table. * Denotes significance at the 5% level; ** at the 1% level.

Table 2 Variables and number of individuals (n) included in the PCAs performed for morphometry (size and shape), plumage melanism (on mantle and wing tip) and coloration of bare parts in the two transects studied (a) in Europe between *L. argentatus* and *L. cachinnans* and (b) in North America between *L. glaucescens* and *L. occidentalis*

Variables included		n	% variance explained by PC1
<i>(a)</i>			
Size	Tarsus, toe, wing, head and bill length, bill depth at gonys and minimum bill depth	140	75
Plumage outer primaries	Length of white and black 10th and 9th primaries, colour of the tongue on the inner web of the 10th primary	148	56 (20 for PC2)
Plumage central primaries	Length of black 8th and 7th primaries, number of primaries with black	148	75
Leg colour	Tarsus, toe and web colour	177	97
Orbital ring colour	Orbital ring colour	140	
Iris colour	Amount of dark spots in the iris	140	
<i>(b)</i>			
Size	Tarsus, toe, culmen, bill and head length; bill depth at gonys, anterior and posterior nares; bill width at posterior nares	311	51
Shape	Tail, wing, 10th, 9th and 8th primary length	311	77
Plumage	Mantle and primary tips colour	392	96
Orbital ring colour	Orbital ring brightness, dominant wavelength and excitation purity	378	82
Bill colour	Bill brightness, dominant wavelength and excitation purity	378	73
Iris colour	Iris brightness, dominant wavelength and excitation purity	378	48
Iris spots colour	IRIS spots brightness, dominant wavelength and excitation purity	378	55

Abbreviation: PCA, principal component analysis.

Percentages of variance explained by the first axis are given in the last column.

Microsatellite analysis

Total genomic DNA from blood samples was extracted using the DNeasy Tissue Kit (Qiagen, Courtaboef, France) following the recommended procedure. Feathers and liver, heart, kidney or pectoral muscle tissues were digested in 10% Chelex 100 (Bio-Rad, Hercules, CA, USA) with 5 μ l of proteinase K followed by two 15-min boiling steps following the procedure described by Walsh *et al.* (1991).

Nine microsatellite loci (HG27, HG25, HG18, HG14, HG16, K31, K32, K67 and K71) were amplified on 178 individuals from the four populations along the European transect, and 273 individuals from 12 populations

along the west coast of North America. Primer sequences and amplification protocols were described elsewhere (Gay *et al.*, 2007). Allele frequencies at the nine microsatellite loci for the populations of the two transects are available in Supplementary Material 1.

We performed probability tests for deviations from the Hardy–Weinberg equilibrium at each microsatellite locus (the results for each locus are presented as Supplementary Material). We estimated the genetic differentiation (F_{ST}) between the most distant populations for each transect (northern Poland and Ukraine in Europe; Alaska populations 1 and 2 vs California populations 18 and 19

in North America—referred to as ‘allopatric F_{ST} ’) for each of the nine microsatellite loci and overall using the software GENETIX (Belkhir *et al.*, 1998). Pairwise F_{ST} values along the two transects were estimated by the parameter θ (Weir and Cockerham, 1984) using GENETIX. We estimated the 95% confidence interval for F_{ST} by bootstrapping over loci. This gives us an estimate of the sampling variance on neutral differentiation. We also estimated the mean number of alleles per locus to ascertain that differentiation did not decrease artificially with the polymorphism of a given locus (Beaumont, 2005).

To check if our nine microsatellite loci provide an accurate assessment of the neutral drift variance among loci, we identified loci potentially affected by selection (outliers) by applying the method of Beaumont and Nichols (1996). The program FDIST2 (<http://www.rubic.rdg.ac.uk/~mab/software.html>) was used to simulate the expected relationship between heterozygosity and F_{ST} for the observed average F_{ST} and assuming selective neutrality. We defined 30 demes and ran 50 000 realizations with a stepwise mutation model. The relationships between observed per locus heterozygosity and F_{ST} were compared to this expected distribution (Luikart *et al.*, 2003).

Phenotypic differentiation

Any evidence for selection uncovered by comparing differentiation between phenotypic traits and molecular markers should be supported by a quantitative genetics design to confirm phenotypic expression. The quantitative genetic differentiation coefficient (Q_{ST}) is defined as follows (Spitze, 1993):

$$Q_{ST} = \frac{V_{A_{between}}}{V_{A_{between}} + 2V_{A_{within}}} \quad (1)$$

where $V_{A_{between}}$ and $V_{A_{within}}$ are the additive genetic variances between and within populations. In practice, Q_{ST} can be measured by quantifying the additive genetic components of variance within and among populations in randomized common garden experiments controlling for non-additive components (for example, using a half-sib design). However, for large white-headed gulls as well as many other organisms, quantitative genetics designs are not practical and there are no long-term pedigree data available (Kruuk, 2004). One solution is to use P_{ST} , the phenotypic analogue of Q_{ST} (Merilä *et al.*, 1997; Storz, 2002; Saint-Laurent *et al.*, 2003; Leinonen *et al.*, 2006; Raeymaekers *et al.*, 2007; Saether *et al.*, 2007)

$$P_{ST} = \frac{V_{P_{between}}}{V_{P_{between}} + 2V_{P_{within}}} \quad (2)$$

P_{ST} can be considered as a proxy for Q_{ST} if we assume that the contribution of environmental and non-additive genetic variance to the differences between populations is low. More generally, if $V_{P_{between}}$ and $V_{P_{within}}$ are the between-population and within-population observed phenotypic variances, and $h^2_{between}$ and h^2_{within} are between-population and within-population additive genetic proportions of differences, respectively, then

$$Q_{ST} = \frac{h^2_{between} V_{P_{between}}}{h^2_{between} V_{P_{between}} + 2h^2_{within} V_{P_{within}}} \quad (3)$$

To avoid artefacts owing to strong environmental effects, we estimated the phenotypic differentiation for a range

of heritability values. The between-population additive genetic proportions of differences ($h^2_{between}$) were allowed to vary between 0 and 1, whereas the within-population proportion (h^2_{within}) equalled 1; 0.75; 0.5 or 0.25. These estimates taking into account the uncertainty in heritability are further called ‘pseudo- Q_{ST} ’. We estimated the phenotypic variances among populations by carrying out a two-factor analysis of variance (ANOVA) (sex and population) using the statistics software R (R-Development-Core-Team, 2004) on the seven synthetic variables (four PC1 scores for Europe (Table 2a): size, melanism of the outer and central primaries, PC2 for outer primaries, tarsus colour and raw data for orbital ring and iris spots; seven PC1 scores for North America (Table 2b): size, shape, plumage melanism, orbital ring, bill, iris and pigment colour). We used the mean square estimates to calculate the between-population ($V_{P_{between}}$) and within-population ($V_{P_{within}}$) variances. We calculated P_{ST} and pseudo- Q_{ST} for each morphological trait, both between the most distant allopatric populations (northern Poland and Ukraine in Europe; Alaska populations 1 and 2 vs California populations 18 and 19 in North America—referred to as ‘allopatric P_{ST} ’) and between all pairwise combinations. The methods available to estimate variance in Q_{ST} estimates (delta method, parametric or non-parametric bootstrap, jackknife, simulation or Bayesian analysis) consider a global Q_{ST} , necessitate a large number of populations and assume an island model. This assumption is not valid in the case of a zone of secondary contact. Instead, we use the 95% confidence interval estimated on F_{ST} by bootstrapping over loci to test whether the estimated Q_{ST} significantly differ from neutral differentiation.

Effect of plasticity on differentiation in allopatry?

In parallel, we investigated the role of environmental variance on phenotypic divergence. First, we examined whether the morphological traits studied varied at the intraspecific level between homo-specific allopatric populations. To detect and exclude individuals with mixed ancestry, microsatellite genotypes of individuals from all populations (four populations in Europe and 12 populations in North America) were assigned to either *L. argentatus* or *L. cachinnans* (in Europe), *L. glaucescens* or *L. occidentalis* (North America) using the Bayesian assignment method (Pritchard *et al.*, 2000) implemented in STRUCTURE (<http://pritch.bsd.uchicago.edu/software.html>). Assignment analyses were performed using microsatellite data, as detailed by Gay *et al.* (2007). Following recommendations by Pritchard *et al.* (2000), individuals were assigned to their putative species, *L. argentatus*, *L. cachinnans*, *L. occidentalis* or *L. glaucescens*, if their assignment probability was higher than 90%. Others were classified as ‘intermediate’. For each transect, we formed two data files containing only ‘pure’ individuals (*L. argentatus* and *L. cachinnans* in Europe; *L. glaucescens* and *L. occidentalis* in North America) and performed ANOVAs (using SAS) on each of the seven synthetic morphological variables (PCA scores). Pairwise *post hoc* Tukey’s tests were performed to check whether there were significant differences between populations (using SAS). Levels of significance were adjusted using sequential Bonferroni corrections (Rice, 1989).

As a second approach to investigate the role of environmental variance, we compared the mean inter-specific differentiation of morphological traits in allopatry and sympatry using a paired *t*-test, where sympatric populations played the role of a common garden (same environmental conditions). If the phenotypic traits analyzed exhibited low plasticity, we expected similar levels of interspecific differentiation in allopatry and sympatry. This comparison was conducted only in the European transect because none of the North American populations had a sufficient number of individuals assigned as both parental species breeding in sympatry.

Evidence for selection on phenotypic traits: F_{ST} – P_{ST} comparison

If the dominant force driving population divergence is genetic drift, the rate of divergence for both microsatellites and quantitative traits is expected to be similar, and the slope of the regression of Q_{ST} against F_{ST} should equal 1. Regression analysis was carried out on pairwise genetic and morphological differentiation estimates for each synthetic variable (PCA scores) with the statistics software SAS version 8 (SAS Institute, 2000). Because a matrix of pairwise comparisons is by construction not independent, we selected a subset of independent values to perform the regression test (using every second value in the diagonal, that is $n/2$ values for n populations). We tested whether the estimated slopes were significantly higher or smaller than 1 ($F_{ST} = P_{ST}$) using a Z-test (Lebreton *et al.*, 1992). This test could only be performed for the North American transect because the European transect did not include enough populations.

Along the North American transect, we tested the effect of latitude on plumage melanism using a generalized linear model in R (R Development Core Team, 2004) with a γ -link function to homogenize the residuals, using the PCA score for the plumage melanism as the dependent variable and latitude, sex and multiple

correspondence analysis score on microsatellite genotypes as explanatory variables (see Gay *et al.*, 2007).

Results

European transect

In the European contact zone, we observed significant deviations from the Hardy–Weinberg equilibrium in all four populations when considering all loci together (Table 1). Looking at each locus separately (see Supplementary Material), the deficit of heterozygote was significant for many loci in the sympatric population (central Poland), as expected after admixture of two differentiated gene pools, but not in northern Poland or southern Poland. In Ukraine, only HG14 locus presented a deficit of heterozygote. Overall, heterozygote deficiency should have a limited effect on the estimates of interpopulation differentiation, as F_{ST} is especially intended to separate deviations from the Hardy–Weinberg equilibrium owing to geographical structuring, as opposed to other causes. The original data for the nine microsatellite loci are available online as Supplementary Material. The genetic differentiation between *L. argentatus* and *L. cachinnans* allopatric populations (northern Poland vs Ukraine, Table 3a) was significant but low ($F_{ST} = 12.5\%$ (0.07–0.17; $P < 10^{-3}$)). To identify loci deviating from equilibrium (potentially under selection), we plotted the expected relationship between heterozygosity at each locus and locus-specific F_{ST} (obtained by simulations) as well as the 95% distribution range and the observed heterozygosity and F_{ST} values for each of the nine loci (Figure 2a). One single locus (K32) fell outside the 0.95 quantile, but it was no longer significant after the Bonferroni correction (Rice, 1989).

Phenotypic differentiation in allopatry for central primaries was high and always exceeded genetic differentiation in neutral markers, even assuming very low levels of additive genetic variance (Figure 3a).

Table 3 Pairwise F_{ST} values based on the nine microsatellite loci for the two transects

(a)												
	Central Poland				Southern Poland				Ukraine			
Northern Poland	0.025				0.147				0.125			
Central Poland					0.067				0.055			
Southern Poland									–0.005 NS			
(b)												
	2	5	6	8	10	11	14	16	17	18	19	
1	0.04	0.06	0.13	0.10	0.18	0.17	0.18	0.22	0.28	0.35	0.33	
2		0.02	0.04	0.03	0.10	0.07	0.09	0.11	0.19	0.24	0.22	
5			0.03	0.01 NS	0.04	0.03	0.04	0.07	0.13	0.17	0.16	
6				0.03	0.03	0.02	0.04	0.04	0.08	0.13	0.12	
8					0.03	0.01 NS	0.01	0.03	0.09	0.14	0.12	
10						0 NS	0 NS	0 NS	0.03	0.05	0.04	
11							–0.01 NS	0.01 NS	0.05	0.08	0.06	
14								0.01 NS	0.05	0.08	0.06	
16									0.03	0.06	0.04	
17										0.01 NS	0.01 NS	
18											0.00 NS	

Four populations in Europe (a) and 12 populations in North America (b) are studied. NS indicates values not significantly different from zero (based on permutation test).

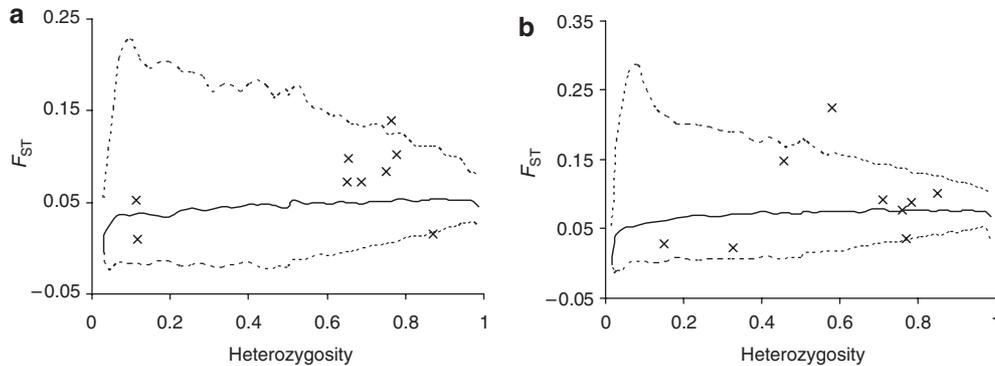


Figure 2 Overall F_{ST} values estimated from nine microsatellite loci (a) between *L. argentatus* and *L. cachinnans* and (b) between *L. glaucescens* and *L. occidentalis* plotted against heterozygosity. Lower, middle and upper lines denote the 2.5, 50, and 97.5 percentiles, respectively, of simulation-based expected distributions assuming an infinite allele mutation model (Beaumont and Nichols, 1996). The simulations were based on F_{ST} representing the average for all microsatellites included.

In contrast, differentiation of leg colour was always lower. The results for the outer primaries (PC2), iris pigmentation and orbital ring colour were more equivocal: these traits showed high levels of differentiation, but only when assuming that the environmental component of phenotypic variance between populations was small (high between-population additive genetic proportions of differences h^2_{between}) (Table 4). As we do not have any data on the heritability of these traits in these populations, it is not possible to compare F_{ST} with P_{ST} for these traits. Nevertheless, *post hoc* Tukey tests on PC1 scores (ANOVA with sex effect) suggest that the variance due to broad-scale environmental differences is limited: only the outer primaries (PC1 only) are significantly variable in *L. cachinnans* (between southern Poland and Ukraine, $P = 0.004$) but not in *L. argentatus*. We also found no significant difference in primary differentiation between *L. argentatus* and *L. cachinnans* in sympatry (central Poland) vs allopatry (northern Poland–Ukraine (Student's test, $t = -0.0232$, $P = 0.982$); however, the central primaries and outer primaries (PC2) did show slightly lower differentiation in sympatry than in allopatry.

North American transect

We observed a significant deficit of heterozygotes in only two populations (6 and 10) close to the centre of the hybrid zone (Table 1), which might be the consequence of sympatry. Most loci showed no or nonsignificant deviations to the Hardy–Weinberg equilibrium (see Supplementary Material 2). In addition, the original data for the nine microsatellite loci is available online as Supplementary Material. Genetic differentiation between allopatric *L. glaucescens* and *L. occidentalis* was high ($F_{ST} = 25.3\%$ (0.162–0.323)), but highly variable depending on the locus considered (variance = 0.023 vs 0.006 in Europe for the same loci). Although most loci exhibited rather high levels of differentiation (around 20–30%), one single locus showed greater structure (K71: $F_{ST} = 44.1\%$) and three other loci were particularly poorly differentiated (<10%). We did not find any significant decrease in the differentiation of a locus with increasing polymorphism, which excludes a Hedrick effect (Hedrick, 1999). Moreover, the locus K71 was identified as an

outlier using the method of Beaumont and Nichols (1996), as illustrated in Figure 2b. The high differentiation at this locus resulted from a steep cline in the frequency of allele 150, one of the rare alleles that reached fixation in *L. glaucescens* populations while its frequency in *L. occidentalis* remained low.

Both plumage melanism and orbital ring colour showed remarkably high differentiation compared with neutral markers, even assuming very low levels of additive genetic variance (Figure 3b and Table 4). On the contrary, differentiation in the iris spots colour was always very low. In addition, the results for some traits were equivocal. The differentiation for iris colour, in particular, was high, but only assuming high between-population additive genetic proportions of differences (h^2_{between}). The results of the ANOVA test for intraspecific differences in phenotypes between populations strengthen the presumption of low plasticity for iris colour, as there was no significant difference between populations of either *L. occidentalis* or *L. glaucescens*. Overall, we found very few significant differences at the intraspecific level. Only two samples of very small size (<4 individuals) yielded significant differences in orbital ring colour and plumage melanism in *L. occidentalis*, and iris pigment and plumage melanism in *L. glaucescens*.

Pairwise estimates of neutral genetic differentiation (pairwise F_{ST}) explained a significant proportion of the pairwise differentiation (P_{ST}) for shape ($R^2 = 72.02\%$; $P = 0.033$), plumage melanism ($R^2 = 69.53\%$; $P = 0.039$), orbital ring ($R^2 = 78.34\%$; $P = 0.019$) and iris colour ($R^2 = 94.13\%$; $P = 0.001$) (Figure 4). The slope of the $P_{ST} = f(F_{ST})$ regression is significantly higher than one ($P_{ST} > F_{ST}$) for orbital ring colour ($P = 0.005$), iris colour ($P = 0.0001$) and plumage melanism ($P = 0.015$). These latter slopes are closer to 2 or 3, indicating that these traits are two or three times more differentiated than neutral markers. For the colour of the iris pigment, the relationship is reversed ($P_{ST} < F_{ST}$) with a slope significantly less than 1 ($P < 0.0001$). It should be noted that these results are based on P_{ST} estimates rather than on pseudo- Q_{ST} . However, seeing that pseudo- Q_{ST} estimates were relatively insensible to variations in heritabilities for orbital ring colour, plumage melanism and colour of the iris pigment (Figure 3), our conclusions are robust.

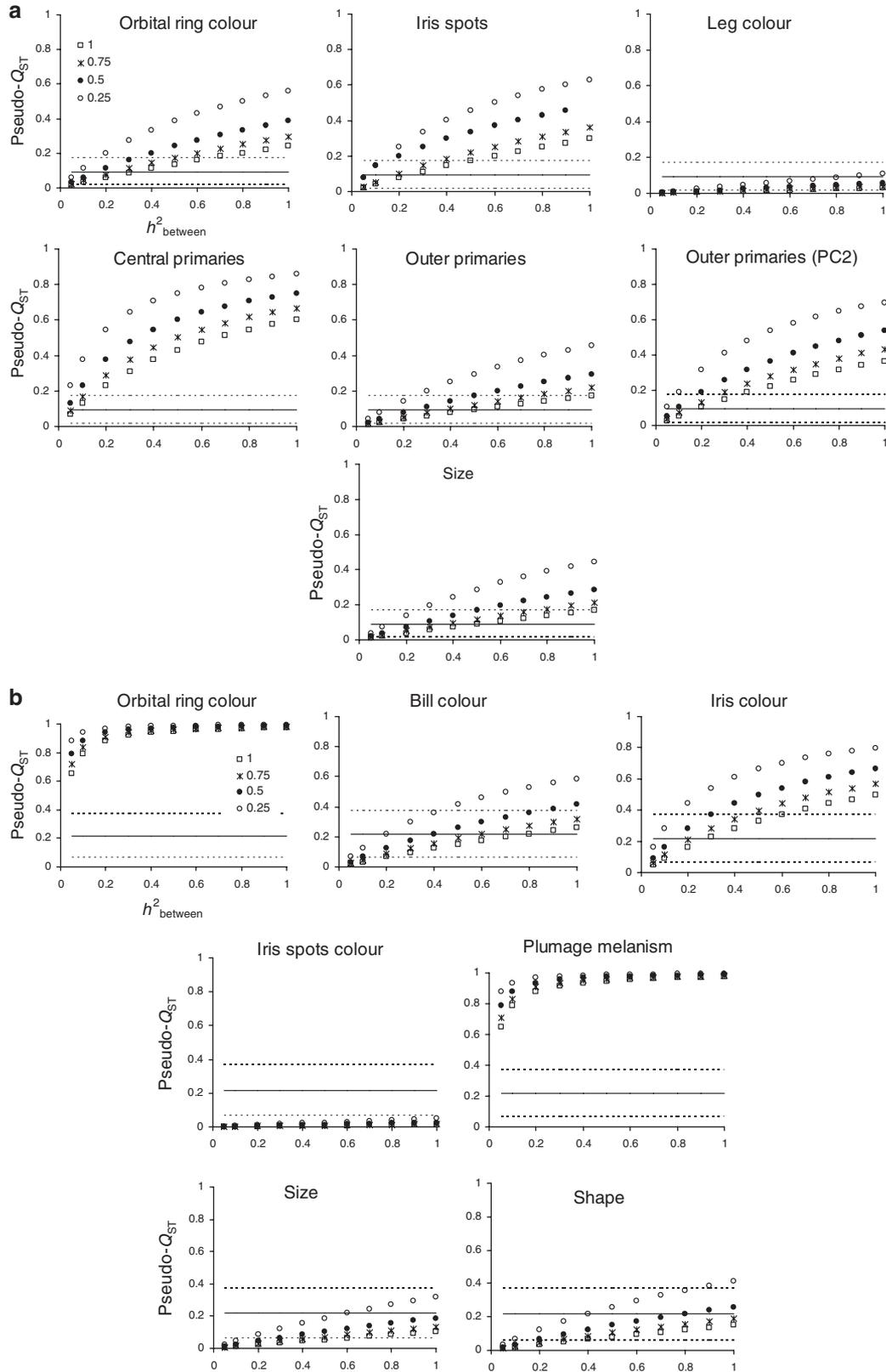


Figure 3 Variation of pseudo- Q_{ST} with the additive genetic proportion of differences between- (x -axis) and within-populations (four different values: 1; 0.75; 0.5 and 0.25). Pseudo- Q_{ST} was estimated for each morphological trait between allopatric populations of (a) *L. argentatus* and *L. cachinnans* (northern Poland vs Ukraine) and (b) *L. glaucescens* and *L. occidentalis* (populations 1 and 2 vs 18 and 19). Estimate of neutral differentiation (average structuration index across nine loci) is represented by the black line \pm 1 s.d. (dotted lines).

Table 4 Phenotypic differentiation in allopatry (P_{ST} , here assuming that the traits are fully heritable) for all synthetic traits summarized by PCA (see Table 2) for the two transects studied

North America		Europe	
	P_{ST}		P_{ST}
Size	0.10	Size	0.17
Shape	0.15	Outer primaries	0.17
Plumage melanism	0.97	Outer primaries (PC2)	0.36
Orbital ring colour	0.97	Central primaries	0.60
Bill colour	0.26	Leg colour	0.03
Iris colour	0.50	Orbital ring colour	0.24
Iris spots colour	0.01	Iris spots	0.30

Abbreviation: PCA, principal component analysis.

P_{ST} was estimated here between the most distant populations in allopatry, between populations 1 and 2 and 18 and 19 in North America ($F_{ST} = 0.253$) and between northern Poland and Ukraine in Europe ($F_{ST} = 0.125$).

Both latitude and genotype at neutral markers (summarized by the multiple correspondence analysis score) as well as their interaction had a significant effect on the plumage melanism (PCA score). This model explained 75% of the variability in plumage melanism among which 17% was explained by latitude, darker plumage being found at lower latitudes.

Discussion

Does morphological differentiation exceed neutral genetic differentiation?

Comparing the genetic differentiation in phenotypic traits and neutral molecular markers provides insights into the relative role of drift and disruptive selection in phenotypic divergence (Merilä and Crnokrak, 2001; McKay and Latta, 2002). In long-lived organisms such as large gulls, it is difficult to have access to the genetic component of phenotypic differentiation unless long-term data are available (Kruuk, 2004). Nevertheless, even assuming a very low heritability, the interspecific divergence between *L. glaucescens* and *L. occidentalis* in plumage melanism and orbital ring colour clearly exceeds neutral differentiation. Similarly, the melanism of the central primaries is highly divergent between *L. argentatus* and *L. cachinnans*. Such a difference cannot arise randomly and must be attributed to spatially varying selection.

How accurately does F_{ST} represent neutral differentiation?

To compare F_{ST} with phenotypic differentiation, it is necessary to ensure that the estimate of neutral genetic variance (F_{ST} on microsatellites) conforms to the expectations of neutral divergence. This is particularly relevant for the North American transect, where we observed a surprisingly high variance in differentiation estimates among microsatellite loci. First, size constraints (homoplasy) could explain this heterogeneity in single locus differentiation. Homoplasy may be most problematic when mutation rates are high, population sizes are large and when there are strong allele size constraints (Estoup *et al.*, 2002; Adams *et al.*, 2004). However, to create such heterogeneity in differentiation, homoplasy should affect

certain loci more than others, for example if the mutation rate varies substantially between different loci.

Second, high levels of within-population variation, as frequently observed for microsatellites, could place an upper limit on the maximal value of F_{ST} (Hedrick, 1999), known as the 'Hedrick effect' (O'Reilly *et al.*, 2004; Larsson *et al.*, 2007). However, this effect occurs primarily when the distribution of genetic variation is largely determined by mutation rather than migration. Dispersal has been shown to be substantial in large gulls: in *L. argentatus*, 35% of birds initiated breeding outside their natal colony (Coulson, 1991) and dispersal distances up to 250 km have been documented (Cramps and Simons, 1983). Moreover, we did not find any clear trend of decreased differentiation with the polymorphism of a locus in either transect. This suggests that the Hedrick effect may not be a major factor here.

Third, variation in the level of differentiation across loci could simply result from sampling variance caused by the finite sample of individuals and alleles (Whitlock and MacCauley, 1999). However, Whitlock and MacCauley (1999) estimated that the sampling variance in F_{ST} decreases with the real F_{ST} value (and their work focused at the intraspecific scale, with F_{ST} estimates close to 1%). We thus expect sampling variance to have only a limited effect on differentiation estimates between *L. glaucescens* and *L. occidentalis*. However, there is also some evidence that the variance in F_{ST} may be inflated in a secondary contact zone depending on the degree of gene flow (Biernie *et al.*, 2003, but see also Murray and Hare, 2006). Without a wider sampling of neutral diversity, it is not possible to exclude sampling variance as an explanation for the high variance in F_{ST} estimates between loci observed here.

Finally, the fourth caveat in using differentiation on molecular markers as an indicator of expected changes under neutral processes is genetic hitchhiking (Maynard Smith and Haigh, 1974). In the *L. glaucescens/L. occidentalis* complex in North America, one single locus (K71) exceeded the mean level of differentiation and was identified as an outlier. We suspect that K71 could be linked to gene(s) undergoing divergent selection. In a secondary contact, we expect a stronger influence of hitchhiking in exacerbating the variance in differentiation because barriers to gene flow (any form of hybrid counter-selection) create linkage disequilibrium between the two populations in contact. The more genes under divergent selection (involved in the barrier to gene flow), the more likely the role of hitchhiking is. Hitchhiking with genes involved in the barrier to gene flow thus increases the variance in rates of introgression across the neutral genome, which further affects the variance in differentiation between loci. In the secondary contact between *L. glaucescens* and *L. occidentalis*, hitchhiking, enhanced by linkage disequilibrium, appears a reasonable explanation for the high variance in differentiation between microsatellite loci.

The high variance observed between loci in F_{ST} estimates emphasizes the need to use a large number of loci to get a decent estimation of neutral differentiation using molecular markers.

How accurately do our P_{ST} estimates reflect the partitioning of genetic variance?

Similar to the Hedrick effect on microsatellites, we could envision that within-population variability could set an

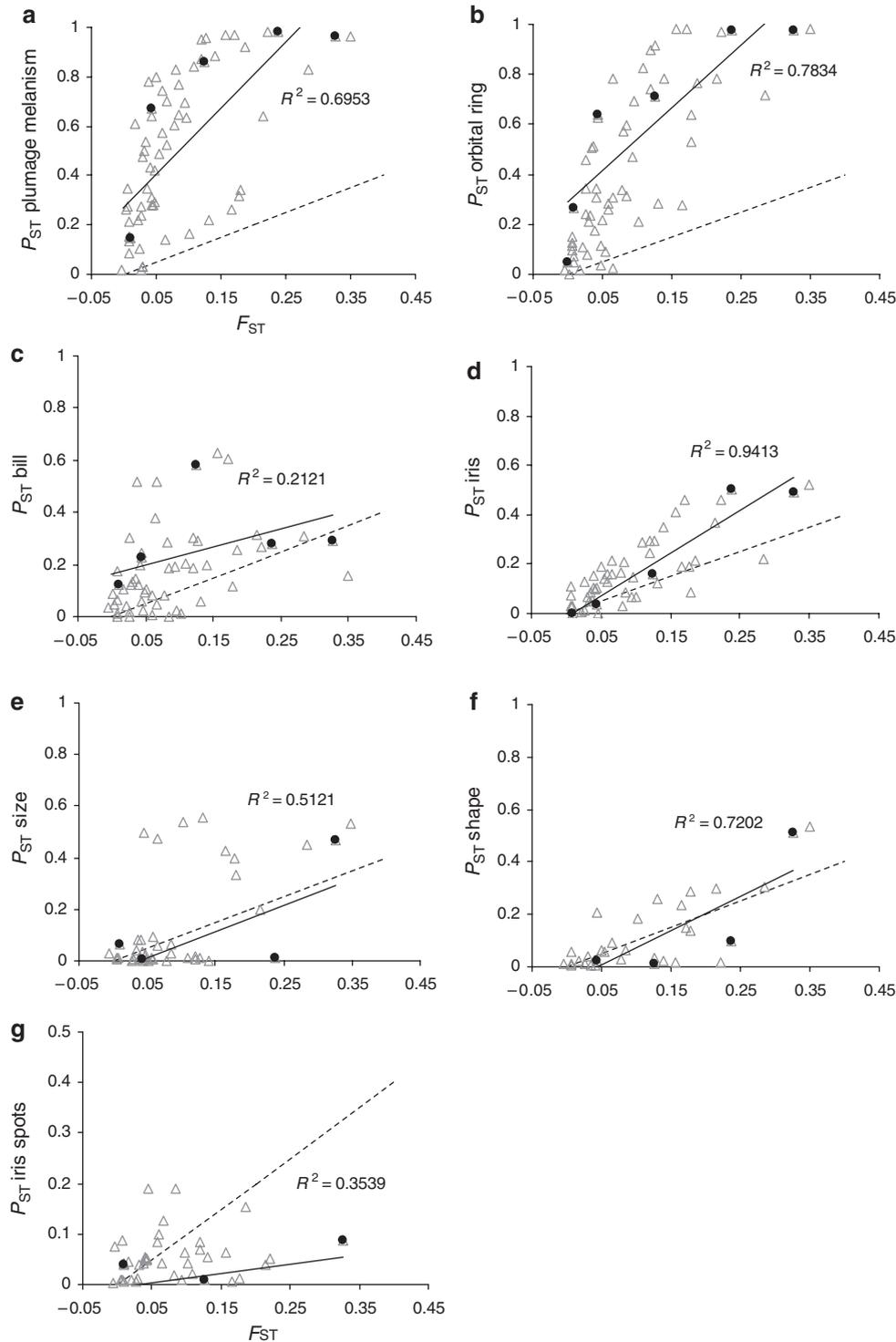


Figure 4 F_{ST} – P_{ST} correlation for (a) plumage melanism, (b) orbital ring colour, (c) bill colour, (d) iris colour, (e) size, (f) shape and (g) iris spots colour between *L. glaucescens* and *L. occidentalis* for all pairwise comparisons between populations (12 populations), assuming complete heritability ($h_{\text{between}}^2 = h_{\text{within}}^2 = 1$). Black dots stand for independent pairwise comparisons, used to test the significance of the correlation and to estimate the slope of the regression. The black line is the regression curve and the dotted line marks the $y = x$ line.

upper limit to the differentiation of quantitative traits. However, this would not affect our conclusion as we found that Q_{ST} estimates were higher than F_{ST} .

Ideally, F_{ST} – Q_{ST} comparisons should be based on purely genetic components of phenotypic variation, removing the environmental component (Spitze, 1993;

Podolsky and Holtsford, 1995; Petit *et al.*, 2001; Edmands and Harrison, 2003; Palo *et al.*, 2003; Gomez-Mestre and Tejedo, 2004; Knopp *et al.*, 2007). Nevertheless, the ‘broad-sense’ Q_{ST} estimates (Leinonen *et al.*, 2008) could be inflated by unaccounted non-additive genetic effects (dominance and epistasis) or persistent maternal

effects (see Merilä and Crnokrak (2001) for a more complete discussion on the assumptions of Q_{ST}). However, maternal effects on plumage melanism are unreported in birds.

In this study, we did not have information on the genetic variance and were thus constrained to use the phenotypic variance between species (P_{ST}), an approach previously taken in a variety of organisms (Rogers and Harpending, 1983; Merilä, 1997; Grapputo *et al.*, 1998; Storz, 2002; Saint-Laurent *et al.*, 2003; Leinonen *et al.*, 2006; Raeymaekers *et al.*, 2007; Saether *et al.*, 2007). The P_{ST} approach helps to detect selection, but assumes that both the environmental variance and the non-additive genetic variance between and within populations are small (Saether *et al.*, 2007). To avoid overestimating population divergence if phenotypic variation is mainly a plastic response to the environment, we analyzed how much our P_{ST} estimates varied with the additive genetic proportion of differences between and within populations (using 'pseudo- Q_{ST} '). We are confident that the high levels of differentiation in plumage melanism and orbital ring colour between *L. glaucescens* and *L. occidentalis* in North America reflect disruptive selection. The melanism of the central primaries in *L. argentatus* and *L. cachinnans* in Europe is also very likely undergoing disruptive selection.

In general, it appears that comparing F_{ST} with Q_{ST} —based on genetic variance—or P_{ST} —based on phenotypic variance—often yields similar patterns (Lynch *et al.*, 1999; Schluter, 2000). There is also no evidence of any systematic bias in P_{ST} studies, as would be expected if phenotypic variation was mainly a plastic response to spatially varying environments (Leinonen *et al.*, 2008).

We found no or low intraspecific phenotypic differentiation (genetic or environmental effects) among populations of each of the four species studied, which suggests that plasticity is not a major component of the between-population variance in the phenotypic traits investigated. The absence of decreased interspecific phenotypic differentiation in sympatry compared with allopatry in the European transect more directly suggests low environmental variance. Yet, some plumage melanism traits showed significant differences at the intraspecific level (in both transects) as well as a tendency of decreased interspecific differentiation in sympatry compared with allopatry (European transect). Plumage melanism was expected to be one of the least plastic traits investigated because its genetic determinism is well established (Theron *et al.*, 2001; Majerus and Mundy, 2003). We suggest that this intraspecific clinal variation in plumage melanism in both transects is shaped by a selection gradient (see below). Overall, plasticity was not sufficient to explain the higher divergence in plumage melanism compared with neutral markers in the two transects studied. Larger sample sizes and ideally controlled breeding experiments are ultimately needed to assess the relative importance of genetically and environmentally based variation in trait values (Storz, 2002).

Nature of the selective pressures

The F_{ST} – P_{ST} comparison provided evidence that selection could have facilitated divergence in plumage melanism in Europe and North America and in orbital ring colour

in North America. Disentangling the selective pressures is, however, a difficult question without a proper experimental design (McDonald *et al.*, 2001) or long-term pedigree data (Przybylo *et al.*, 2000; Merilä *et al.*, 2001; Hosken and Balloux, 2002). Indirect approaches, such as regression models, can reveal some of the environmental factors (habitat, diet, and so on) potentially involved in the phenotypic divergence (Storz, 2002; Palo *et al.*, 2003).

According to Gloger's rule, sun and humidity could generate strong selective pressures for more resistant, darker melanin-rich feathers (Hayes, 2001) because sun and humidity promote bacterial degradation of feathers and damage birds' plumage (Burt and Ichida, 2004). Other hypotheses state that the damage of heat stress could be reduced in dark-coloured species (Ward *et al.*, 2002). In gulls, melanin pigments are responsible for the intensity of light grey to black in the feathers of the mantle and wing tips. Species living at lower latitudes would thus be more likely to exhibit darker plumage than similar species living at northern latitudes. This is compatible with the observed pattern of darker plumage in southern *L. occidentalis* in North America and *L. cachinnans* in Europe vs their more northerly counterparts. Clinal variation of plumage traits observed at the intraspecific scale (Bell, 1997) also matches the selective gradient of sun and humidity. We found that latitude explains another 17% of variation in plumage melanism in addition to the effect of drift expressed as the variation in neutral markers. Melanin pigments are also deposited in the iris (dark spots). This trait measured in *L. argentatus* and *L. cachinnans* similarly formed a clear latitudinal gradient from *L. argentatus* to *L. cachinnans*. Both melanin-based traits thus support Gloger's rule, in Europe as well as in North America.

The second trait for which we found evidence of selection is the carotenoid-based orbital ring colour in North America. The amount of carotenoid pigments available is known to vary with diet and body condition (Blount, 2004), which could generate high within-population phenotypic variance. However, we found no significant difference between populations for either species. It was shown in a closely related species (*Larus marinus*) that the intensity of carotenoid coloration of the bare parts is an honest signal of individual quality (Kristiansen *et al.*, 2006) and that the coloration of bare parts, especially bill and foot coloration, may be involved in mate choice in seabirds (Pierotti, 1987). Bell (1997) found significant assortative mating between gulls in mixed colonies of *L. glaucescens*, *L. occidentalis* and their hybrids. We suggest that the orbital ring colour could have diverged through sexual selection, if females of each taxa exhibit different preferences for their partner's orbital ring colour. In sympatry, such disruptive selection resulting from assortative mating could maintain the differentiation despite interspecific gene flow.

Speciation with gene flow

As it is generally the case for most birds, species in the large white-headed gull complex probably initially diverged in allopatry (Liebers *et al.*, 2004). A previous

study suggested that the interspecific gene flow is involved in the low level of genetic differentiation between the recently diverged large white-headed gull species, based on comparisons of mitochondrial and nuclear markers as well as geographic arguments (Crochet *et al.*, 2003). However, even very little gene flow between populations is sufficient to maintain genetic equanimity (Wright, 1931). Only selection can override the effects of gene flow on some traits and maintain species distinctness (Ehrlich and Raven, 1969). Thus, closely related taxa in which phenotypic cohesion and mate recognition are caused by a small number of loci are expected to share much of their genetic variation. This pattern has frequently been observed in birds despite highly structured phenotypic variability (Grapputo *et al.*, 1998; Freeland and Boag, 1999; Podos, 2001; Clegg *et al.*, 2002; Hosken and Balloux, 2002; Chan and Arcese, 2003). Our comparisons of differentiation levels in phenotypic traits and neutral markers confirmed that plumage melanism and orbital ring colour likely diverged through disruptive selection. Even if evidence of disruptive selection is not sufficient to conclude that there is speciation with gene flow, the absence of disruptive selection would have definitely ruled out this scenario. Our results thus substantiate a scenario of speciation with gene flow for the large white-headed gulls. More generally, some of the recently diverged species that hybridize when coming into secondary contact might actually only differ in restricted regions of the genome that, by virtue of their phenotypic effects, are undergoing strong disruptive selection and/or are subject to assortative mating. This emphasizes the benefit of multilocus approaches in a population genetics framework for studies of speciation, especially to assess the role of gene flow and natural selection in species divergence.

Acknowledgements

We thank AI and VA Koshelev and the polish teams 'Czaplon' and 'Kuling', who helped with fieldwork. In particular, we thank M Faber, J Betleja, J Pietrasik, P Ziecik, A Forsten, V Rauste, A Lindholm, R Barret and R Bradburry for collecting samples and taking morphological measurements. Fieldwork in Poland was supported by the Polish Committee for Scientific Research Grant 6PO4C 04719 and the PAI Polonium Grant 5572.I/2004. We are grateful to C Cicero (Museum of Vertebrate Zoology, University of California, Berkeley, CA, USA) who sent us the samples for the North American hybrid zone. We also thank P Sourrouille and B Nabholz in Montpellier for technical assistance with the molecular work. J Merilä, DJ Hosken and M Demont as well as three anonymous reviewers made helpful comments on previous versions of this manuscript.

References

Adams RI, Brown KM, Hamilton MB (2004). The impact of microsatellite electromorph size homoplasy on multilocus population structure estimates in a tropical tree (*Corythophora alta*) and an anadromous fish (*Morone saxatilis*). *Mol Ecol* **13**: 2579–2588.

Beaumont MA (2005). Adaptation and speciation: what can F_{ST} tell us? *Trends Ecol Evol* **20**: 435–440.

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

Belkhir K, Borsa P, Goudet J, Chikhi L, Bonhomme F (1998). *GENETIX, logiciel sous Windows TM pour la génétique des populations*. Laboratoire Génome et Populations, CNRS UPR 9060, Université de Montpellier II: Montpellier, France.

Bell DA (1996). Genetic differentiation, geographic variation and hybridization in gulls of the *Larus glaucescens-occidentalis* complex. *Condor* **98**: 527–546.

Bell DA (1997). Hybridization and reproductive performance in gulls of the *Larus glaucescens-occidentalis* complex. *Condor* **99**: 585–594.

Bierne N, Daguin C, Bonhomme F, David P, Borsa P (2003). Direct selection on allozymes is not required to explain heterogeneity among marker loci across a *Mytilus* hybrid zone. *Mol Ecol* **12**: 2505–2510.

Blount JD (2004). Carotenoids and life-history evolution in animals. *Arch Biochem Biophys* **430**: 10–15.

Burt EH, Ichida JM (2004). Gloger's rule, feather-degrading bacteria, and color variation among song sparrows. *Condor* **106**: 681–686.

Chan Y, Arcese P (2003). Morphological and microsatellite differentiation in *Melospiza melodia* (Aves) at a microgeographic scale. *J Evol Biol* **16**: 939–947.

Clegg SM, Degnan SM, Moritz C, Estoup A, Kikkawa J, Owens IPF (2002). Microevolution in island forms: the roles of drift and directional selection in morphological divergence of a passerine bird. *Evolution* **56**: 2090–2099.

Coulson JC (1991). The population dynamics of culling herring gulls and lesser black-backed gulls. In: Perrins CM, Lebreton JD, Hiron GJM (eds). *Bird Population Studies*. Oxford University Press: Oxford.

Cramps S, Simmons KEL (1983). *Handbook of the Birds of Europe the Middle East and North Africa. The Birds of the Western Palearctic*. Oxford University Press: Oxford, UK.

Crochet PA, Chen JJZ, Pons JM, Lebreton JD, Hebert PDN, Bonhomme F (2003). Genetic differentiation at nuclear and mitochondrial loci among large white-headed gulls: sex-biased interspecific gene flow? *Evolution* **57**: 2865–2878.

Doebeli M, Dieckmann U, Metz JAJ, Tautz D (2005). What we have also learned: adaptive speciation is theoretically plausible. *Evolution* **59**: 691–695.

Edmands S, Harrison JS (2003). Molecular and quantitative trait variation within and among populations of the intertidal copepod *Tigriopus californicus*. *Evolution* **57**: 2277–2285.

Ehrlich PR, Raven PH (1969). Differentiation of populations. *Science* **165**: 1228–1232.

Estoup A, Jarne P, Cornuet JM (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol* **11**: 1591–1604.

Faber M, Betleja J, Malczyk P, Gwiazda R (2001). Mixed colonies of large white-headed gulls in southern Poland. *Br Birds* **94**: 529–534.

Freeland JR, Boag PT (1999). The mitochondrial and nuclear genetic homogeneity of the phenotypically diverse Darwin's ground finches. *Evolution* **53**: 1553–1563.

Gavrilets S (2003). Perspective: models of speciation: what have we learned in 40 years? *Evolution* **57**: 2197–2215.

Gay L, Neubauer G, Zagalska-Neubauer M, Debain C, Pons JM, David P *et al.* (2007). Molecular and morphological patterns of introgression between two large white-headed gull species in a zone of recent secondary contact. *Mol Ecol* **16**: 3215–3227.

Gomez-Mestre I, Tejedo M (2004). Contrasting patterns of quantitative and neutral genetic variation in locally adapted populations of the natterjack toad, *Bufo calamita*. *Evolution* **58**: 2343–2352.

- Good TP, Ellis JC, Annett CA, Pierotti R (2000). Bounded hybrid superiority in an avian hybrid zone: effects of mate, diet, and habitat choice. *Evolution* **54**: 1774–1783.
- Grapputo A, Pilastro A, Marin G (1998). Genetic variation and bill size dimorphism in a passerine bird, the reed bunting *Emberiza schoeniclus*. *Mol Ecol* **7**: 1173–1182.
- Hayes FE (2001). Geographic variation, hybridization, and the leapfrog pattern of evolution in the Suiriri flycatcher (*Suiriri suiriri*) complex. *Auk* **118**: 457–471.
- Hedrick PW (1999). Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**: 313–318.
- Hosken DJ, Balloux F (2002). Thirty years of evolution in Darwin's finches. *Trends Ecol Evol* **17**: 447–448.
- Johnson PA, Gullberg U (1998). Theory and models of sympatric speciation. In: Howard DJ, Berlocher SH (eds). *Endless Forms: Species and Speciation*. Oxford University Press: Oxford, UK. pp 79–89.
- Kirkpatrick M, Ravigne V (2002). Speciation by natural and sexual selection: models and experiments. *Am Nat* **159**: S22–S35.
- Knopp T, Cano JM, Crochet PA, Merilä J (2007). Contrasting levels of variation in neutral and quantitative genetic loci on island populations of moor frogs (*Rana arvalis*). *Conserv Genet* **8**: 45–56.
- Kristiansen KO, Bustnes JO, Folstad I, Helberg M (2006). Carotenoid coloration in great black-backed gull *Larus marinus* reflects individual quality. *J Avian Biol* **37**: 6–12.
- Kruuk LEB (2004). Estimating genetic parameters in natural populations using the 'animal model'. *Philos Trans R Soc Lond B Biol Sci* **359**: 873–890.
- Lande R (1992). Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. *Evolution* **46**: 381–389.
- Larsson LC, Laikre L, Palm S, Andre C, Carvalho GR, Ryman N (2007). Concordance of allozyme and microsatellite differentiation in a marine fish, but evidence of selection at a microsatellite locus. *Mol Ecol* **16**: 1135–1147.
- Lebreton JD, Burnham KP, Clobert J, Anderson DR (1992). Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecol Monogr* **62**: 67–118.
- Leinonen T, Cano JM, Mäkinen H, Merilä J (2006). Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. *J Evol Biol* **19**: 1803–1812.
- Leinonen T, O'Hara RB, Cano JM, Merilä J (2008). Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *J Evol Biol* **21**: 1–17.
- Lessells CM, Boag PT (1987). Unrepeatable repeatabilities—a common mistake. *Auk* **104**: 116–121.
- Liebers D, de Knijff P, Helbig AJ (2004). The herring gull complex is not a ring species. *Proc R Soc Lond B Biol Sci* **271**: 893–901.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003). The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* **4**: 981–994.
- Lynch M, Pfrender M, Spitze K, Lehman N, Hicks J, Allen D et al. (1999). The quantitative and molecular genetic architecture of a subdivided species. *Evolution* **53**: 100–110.
- Machado CA, Kliman RM, Markert JA, Hey J (2002). Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Mol Biol Evol* **19**: 472–488.
- Majerus MEN, Mundy NI (2003). Mammalian melanism: natural selection in black and white. *Trends Genet* **19**: 585–588.
- Maynard Smith J, Haigh J (1974). The hitch-hiking effect of a favourable gene. *Genet Res* **23**: 23–35.
- McDonald DB, Clay RP, Brumfield RT, Braun MJ (2001). Sexual selection on plumage and behavior in an avian hybrid zone: experimental tests of male–male interactions. *Evolution* **55**: 1443–1451.
- McKay JK, Latta RG (2002). Adaptive population divergence: markers, QTL and traits. *Trends Ecol Evol* **17**: 285–291.
- Merilä J (1997). Quantitative trait and allozyme divergence in the Greenfinch (*Carduelis chloris*, Aves: Fringillidae). *Biol J Linn Soc Lond* **61**: 243–266.
- Merilä J, Björklund M, Baker AJ (1997). Historical demography and present day population structure of the Greenfinch, *Carduelis chloris*—an analysis of mtDNA control-region sequences. *Evolution* **51**: 946–956.
- Merilä J, Crnokrak P (2001). Comparison of genetic differentiation at marker loci and quantitative traits. *J Evol Biol* **14**: 892–903.
- Merilä J, Kruuk LEB, Sheldon BC (2001). Natural selection on the genetical component of variance in body condition in a wild bird population. *J Evol Biol* **14**: 918–929.
- Munsell AH (1968). *Dominant Wavelength and Excitation Purity for Designated Munsell Color Notations*. Munsell Color Co., Macbeth Kollmorgen: Baltimore.
- Munsell AH (1971). *Munsell Neutral Value Scale. Glossy. 37-Step Scale*. Munsell Color Co., Macbeth Kollmorgen: Baltimore.
- Munsell AH (1976). *Munsell Book of Color*. Munsell Color Co., Macbeth Kollmorgen: Baltimore.
- Murray MC, Hare MP (2006). A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*. *Mol Ecol* **15**: 4229–4242.
- Neubauer G, Zagalska-Neubauer M, Gwiazda R, Faber M, Bukacinski D, Betleja J et al. (2006). Breeding large gulls in Poland: distribution, numbers, trends and hybridization. *Vogelwelt* **127**: 11–22.
- Noor MAF, Garfield DA, Schaeffer SW, Machado CA (2007). Divergence between the *Drosophila pseudoobscura* and *D. persimilis* genome sequences in relation to chromosomal inversions. *Genetics* **177**: 1417–1428.
- O'Reilly PT, Canino MF, Bailey KM, Bentzen P (2004). Inverse relationship between F_{ST} and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. *Mol Ecol* **13**: 1799–1814.
- Olsen KM, Larsson H (2004). *Gulls of Europe, Asia and North America*. A & C Black Publishers Ltd.: London.
- Palo JU, O'Hara RB, Laugen AT, Laurila A, Primmer CR, Merilä J (2003). Latitudinal divergence of common frog (*Rana temporaria*) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data. *Mol Ecol* **12**: 1963–1978.
- Panov EN, Monzиков DG (1999). Intergradation between the herring gull *Larus argentatus* and the southern herring gull *Larus cachinnans* in European Russia. *Russ J Zool* **3**: 129–141.
- Petit C, Freville H, Mignot A, Colas B, Riba M, Imbert E et al. (2001). Gene flow and local adaptation in two endemic plant species. *Biol Conserv* **100**: 21–34.
- Pierotti R (1987). Isolating mechanisms in seabirds. *Evolution* **41**: 559–570.
- Podolsky RH, Holtsford TP (1995). Population structure of morphological traits in *Clarkia dudleyana*. 1. Comparison of F_{ST} between allozymes and morphological traits. *Genetics* **140**: 733–744.
- Podos J (2001). Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature* **409**: 185–188.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Przybylo R, Sheldon BC, Merilä J (2000). Patterns of natural selection on morphology of male and female collared flycatchers (*Ficedula albicollis*). *Biol J Linn Soc Lond* **69**: 213–232.

- R Development Core Team (2004). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria.
- Raeymaekers JAM, Van Houdt JIJ, Larmuseau MHD, Geldof S, Volckaert FAM (2007). Divergent selection as revealed by P_{ST} and QTL-based F_{ST} in three-spined stickleback (*Gasterosteus aculeatus*) populations along a coastal-inland gradient. *Mol Ecol* **16**: 891–905.
- Reed DH, Frankham R (2001). How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* **55**: 1095–1103.
- Rice WR (1989). Analyzing tables of statistical tests. *Evolution* **43**: 223–225.
- Rogers AR, Harpending HC (1983). Population structure and quantitative characters. *Genetics* **105**: 985–1002.
- Saether SA, Fiske P, Kalas JA, Kuresoo A, Luigujoe L, Piertney SB et al. (2007). Inferring local adaptation from Q_{ST} – F_{ST} comparisons: neutral genetic and quantitative trait variation in European populations of great snipe. *J Evol Biol* **20**: 1563–1576.
- Saint-Laurent R, Legault M, Bernatchez L (2003). Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (*Osmerus mordax Mitchill*). *Mol Ecol* **12**: 315–330.
- Sanou H, Lovett PN, Bouvet JM (2005). Comparison of quantitative and molecular variation in agroforestry populations of the shea tree (*Vitellaria paradoxa* CF Gaertn) in Mali. *Mol Ecol* **14**: 2601–2610.
- SAS Institute (2000). *SAS/STAT Software: Version 8*. SAS Institute, Inc.: Cary, NC.
- Schluter D (2000). *The Ecology of Adaptive Radiation*. Oxford University Press: Oxford.
- Schluter D (2001). Ecology and the origin of species. *Trends Ecol Evol* **16**: 372.
- Spitze K (1993). Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics* **135**: 367–374.
- Steinger T, Haldimann P, Leiss KA, Muller-Scharer H (2002). Does natural selection promote population divergence? A comparative analysis of population structure using AFLP markers and quantitative traits. *Mol Ecol* **11**: 2583–2590.
- Storz JF (2002). Contrasting patterns of divergence in quantitative traits and neutral DNA markers: analysis of clinal variation. *Mol Ecol* **11**: 2537–2551.
- Theron E, Hawkins K, Bermingham E, Ricklefs RE, Mundy NI (2001). The molecular basis of an avian plumage polymorphism in the wild: a melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, *Coereba flaveola*. *Curr Biol* **11**: 550–557.
- Thioulouse J, Chessel D, Dolédec S, Olivier JM (1997). ADE-4: a multivariate analysis and graphical display software. *Stat Comput* **7**: 75–83.
- Via S (2001). Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol Evol* **16**: 381–390.
- Via S (2002). The ecological genetics of speciation. *Am Nat* **159**: S1–S7.
- Volis S, Yakubov B, Shulgina I, Ward D, Mendlinger S (2005). Distinguishing adaptive from nonadaptive genetic differentiation: comparison of Q_{ST} and F_{ST} at two spatial scales. *Heredity* **95**: 466.
- Walsh PS, Metzger DA, Higuchi R (1991). Chelex-100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10**: 506–513.
- Wang RL, Wakeley J, Hey J (1997). Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* **147**: 1091–1106.
- Ward JM, Blount JD, Ruxton GD, Houston DC (2002). The adaptive significance of dark plumage for birds in desert environments. *Ardea* **90**: 311–323.
- Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
- Whitlock MC, MacCauley DE (1999). Indirect measures of gene flow and migration: F_{ST} not equal $1/(1+4Nm)$. *Heredity* **82**: 117–125.
- Wright S (1931). Evolution in Mendelian populations. *Genetics* **16**: 97–159.
- Wu C-I (2001). The genic view of the process of speciation. *J Evol Biol* **14**: 851–865.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)