

ORIGINAL ARTICLE

Using principle component analysis to compare genetic diversity across polyploidy levels within plant complexes: an example from British Restharrow (*Ononis spinosa* and *Ononis repens*)

JM Kloda¹, PDG Dean², C Maddren¹, DW MacDonald¹ and S Mayes³

¹Department of Genetics, University of Cambridge, Cambridge, UK; ²Cambio Ltd, The Irwin Centre, Dry Drayton, Cambridge, UK and ³Division of Agricultural and Environmental Sciences, School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, Leicestershire, UK

The investigation of genetic diversity between related plant populations which differ in ploidy levels is problematic, with common statistical methods developed for diploids being inappropriate for polyploid species. Studies into gene flow in such complexes are critical and can shed light on the mechanisms that generate and maintain populations of different polyploidy levels. We have investigated the use of principle component (PCO) analysis as one approach to elucidate population structure within British Restharrow (*Leguminosae*, *Ononis* spp). Restharrow were common agricultural weed species until the advent of mechanical ploughing and both diploid ($2n=2x=30$; *O. spinosa* and *O. intermedia*) and tetraploid ($2n=4x=60$; *O. repens* and *O. maritima*) taxa exist. Patterns of genetic diversity were investigated among British Restharrow using 10 microsatellite loci with 21 Restharrow populations analysed (411

individual plants) from Central and Eastern Britain. PCO analysis revealed clear genetic differentiation of the sampled plants into two groups, corresponding to *O. spinosa/O. intermedia* (diploid) and *O. repens/O. maritima* (tetraploid) plants. Evidence of genetic differentiation by distance was also revealed for *O. repens/O. maritima*, but not for *O. spinosa/O. intermedia*. The data suggest the presence of strong reproductive barriers between diploid and tetraploid Restharrow in Britain, but not within ploidy levels. This genetic isolation between ploidy levels is confirmed by a detailed analysis of a sympatric site (Harton Down Hill). These results demonstrate that PCA analysis is a suitable general tool for comparing related species of different ploidy levels.

Heredity (2008) **100**, 253–260; doi:10.1038/sj.hdy.6801044; published online 12 December 2007

Keywords: *Ononis*; microsatellite; SSR; phylogeography; polyploidy; gene flow

Introduction

Many plants, such as Restharrow (*Fabaceae*, *Leguminosae*) have ploidy level differences between species within the same complex. Whether gene flow still occurs between potential sibling species is an important question whose answer may have fundamental implications for the population structure present in those species. However, comparisons of genetic structure and relatedness across ploidy levels are complicated by the underlying assumption of disomic inheritance in most methods of analysis (for example, Wright's *F*_{st}; Wright, 1943, 1946, 1951). While the population genetics of autotetraploids have been modelled and elucidated (Glendinning, 1989; Moody *et al.*, 1993; Ronfort *et al.*, 1998), the theory is not readily put into practice due to the technical problem of not being able to assign copy number

to the molecular markers studied, even using highly robust locus-specific marker systems such as microsatellites (Ellegren *et al.*, 1995; Schlotterer, 1998; Amos, 1999; Chambers and MacAvoy, 2000). In this paper, we test the value of principle component (PCO) analysis of microsatellite data for making an assessment of such genetic relationships, using British Restharrow as an example.

Sixty-seven species of *Ononis* were described by Sirjaev (1932) with one new species being added by Crespo and Serra (1993). The genus is widely distributed throughout Europe, North East Africa and North America. Many *Ononis* species also produce a waxy substance in their tap roots called Onocerin (Rowan and Dean, 1972), which has been used as one criterion for taxonomic classification. In the British Isles, there are between two and five taxa (depending upon classification), which occur as native populations. *O. reclinata* ($2n=4x=60$) is a native (but very rare taxon) confined to just three populations (Wigginton, 1999). It is not included in this current analysis. The remaining taxa have a wide distribution, from coastal dunes and shingle beaches to grassy meadows.

Before this study, there had been no DNA-based evidence published on the genetic relationships between

Correspondence: Dr S Mayes, Division of Agricultural and Environmental Sciences, School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, Leicestershire, LE12 5RD, UK.

E-mail: sean.mayes@nottingham.ac.uk

Received 3 August 2006; revised 25 June 2007; accepted 5 July 2007; published online 12 December 2007

Restharrow in Britain, with the classification to date being based upon: gross morphology; multivariate analysis of morphological characters; analysis of stomatal structure and seed-coat ornamentation; comparative root anatomy; chromosome number; floral morphology; terminal leaflet morphology; flavonoid chemistry; palynology; α -onocerin and sterol content; observation of and experiments in hybridization (Table 1; Sirjaev, 1932; Morton, 1956; Morisset, 1967; Ivimey-Cook, 1968; Rowan and Dean, 1972; Stephens, 1978, 1979a,b; Sañudo *et al.*, 1979; Gupta, 1980; Cannon, 1981; Saint-Martin, 1992; Langer *et al.*, 1995; Rodriguez-Riano *et al.*, 1999).

Despite this effort, the taxonomy of the main British Restharrow species remains controversial. Those found most commonly in Britain (that is, excluding *O. reclinata*) are variously classified into between one and four species, according to different authors (Table 1). This long-running debate has so far failed to reach a consensus, and has led to a lack of taxonomic clarity. For the current study, we will designate the taxa as: *O. repens* ($2n = 4x = 60$); *O. maritima* ($2n = 4x = 60$); *O. spinosa* ($2n = 2x = 30$); *O. intermedia* ($2n = 2x = 30$). We have chosen this approach for clarity, given the lack of taxonomic guidance available and will re-evaluate these designations in the Discussion in the light of the understanding developed in this study.

Meiotic behaviour had also been studied in an attempt to clarify relationships among diploid and tetraploid Restharrow. The chromosomes of tetraploid *O. repens*

and *O. maritima* are known to produce multivalents, although the difficulty of obtaining good meiotic spreads prevented a quantitative analysis (Morisset, 1967, 1978). This would argue for an autotetraploid origin. One possibility would be the autotetraploidization of *O. spinosa* or *O. intermedia*, which would imply the potential for significant unidirectional gene flow from diploid to tetraploid Restharrow.

Masterson (1994) estimated that approximately 70% of angiosperms have polyploidy in their evolutionary history, so an improved understanding of the processes by which polyploid species arise and are maintained is crucial for the understanding of evolution of angiosperms. Few molecular genetic studies spanning diploids and their corresponding polyploids have been published. Where they have been, a number of approaches have been adopted (for example, Lumaret and Barrientos, 1990; Gauthier *et al.*, 1998; Raker and Spooner, 2002; Robertson *et al.*, 2004). In this study, microsatellites data are analysed using PCO analysis to investigate the relationships amongst British Restharrow, to identify whether gene flow has occurred in the recent past between the predominantly diploid and tetraploid Restharrow taxa and also between the taxa with the same ploidy levels. We also evaluate whether PCO analysis coupled to microsatellite analysis could be used to overcome some of the limitations that exist with conventional disomic analysis tools, when the species of interest include polyploids.

Table 1 Distinguishing characteristics and previous taxonomic treatment of common Restharrow species in the British Isles

Description	<i>O. spinosa</i>	<i>O. intermedia</i>	<i>O. repens</i>	<i>O. maritima</i>
Ploidy levels	Diploid ($2n = 2x = 30$)	Primarily diploid ($2n = 2x = 30$)		Tetraploid ($2n = 4x = 60$)
Distinguishing features	Spiny, nonrhizomatous, with upright habit, red woody stems and irregular hairs.	Intermediate features between <i>O. spinosa</i> and <i>O. repens</i>	Occasionally spiny, rhizomatous, decumbent, rarely ascending with hairy stems. More glandular than the diploids	Adopts a flatter posture than <i>O. repens</i>
Common features	Perennial shrub with pink solitary flowers in lax racemes		and unifoliate (although occasionally trifoliate) leaves	
Distribution	Well distributed in the British Isles, noncoastal		Largely restricted primarily to England, with <i>O. repens</i> being inland and <i>O. maritima</i> being coastal on shingle beaches and cliff-tops	
Soils	Preference for heavy clay limestone soils		Preference for light, well drained soils	
Author				
Linnaeus (1753); Morisset (1967)	<i>O. spinosa</i> L.		<i>O. repens</i> L.	
Greuter (1986); Sell and Murrell (2006)	<i>O. spinosa</i> subsp. <i>spinosa</i>	<i>O. spinosa</i> subsp. <i>intermedia</i> (Rouy) P. Fourn.	<i>O. spinosa</i> subsp. <i>procurrens</i> (Wallr.) Briq.	<i>O. spinosa</i> subsp. <i>maritima</i> (Dumort.) P. Fourn.
Jarvis <i>et al.</i> (1983); Cannon (1981); Ivimey-Cook (1968)	<i>O. campestris</i> G. Koch		<i>O. repens</i> L.	
Sirjaev (1932)	<i>O. spinosa</i> L. (divided into 4 subspecies plus 11 varieties)		<i>O. repens</i> L. (divided into 8 varieties)	
Stace (1997)	<i>O. spinosa</i> L.	<i>O. x pseudohircina</i> Schur.	<i>O. repens</i> L.	<i>O. repens</i> subsp. <i>maritima</i>
Rouy and Foucaud (1897)	<i>O. campestris</i> Koch et Ziz	<i>O. intermedia</i> Mey	<i>O. procurrens</i> Wallr	<i>O. repens</i> L.

Table 2 Sampling sites (coordinates) of British Restharrow populations used in this study

Sample	Location	Coordinates (longitude)	Coordinates (latitude)	Species by appearance	Number of samples
1	Aldeburgh, Suffolk	1:36:19	52:09:36	<i>O. maritima</i>	24
2	Ascott Under Wychwood, Oxfordshire	1:33:09	51:51:40	<i>O. repens</i>	24
3	Blackhall Cliff, Durham	1:16:08	54:44:42	<i>O. intermedia</i>	24
4	Blackhall Cliff Base, Durham	1:16:08	54:44:42	<i>O. maritima</i>	24
5	Bishops Middleham, Durham	1:29:14	54:41:19	<i>O. repens</i>	24
6	Cheesefoot Head, Hampshire	1:15:08	51:03:09	<i>O. repens</i>	24
7	Cleadon, Tyne and Wear	1:23:43	54:57:51	<i>O. spinosa</i>	24
8	Dunwich, Suffolk	1:38:04	52:16:50	<i>O. maritima</i>	24
9	Garforth, West Yorkshire	1:21:55	53:46:42	<i>O. intermedia</i>	24
10	Gog Magog Golf Club, Cambridgeshire	0:10:36	52:09:56	<i>O. repens</i>	14
12s	Marsden, Tyne and Wear	1:23:32	54:58:57	<i>O. spinosa</i>	20
12r	Marsden, Tyne and Wear	1:23:32	54:58:57	<i>O. repens</i>	7
13i	Marsden Rocks, Tyne and Wear	1:22:08	54:58:24	<i>O. intermedia</i>	12
14	Old Hall Marshes, Essex	0:50:04	51:47:07	<i>O. spinosa</i>	8
15	Orwell field, Cambridgeshire	0:00:31	52:08:16	<i>O. spinosa</i>	24
16	Orwell Foxhill, Cambridgeshire	0:00:13	52:08:32	<i>O. spinosa</i>	24
17	Pelsall, Clayhanger, West Midlands	1:56:53	52:38:34	<i>O. repens</i>	16
18	Purgatory, Oxfordshire	1:20:17	51:53:53	<i>O. repens</i>	8
19	Ryton Willows, Tyne and Wear	1:45:22	54:58:52	<i>O. repens</i>	7
21	St Neots, Cambridgeshire	0:16:19	52:14:11	<i>O. spinosa</i>	23
22	Thorpeness, Suffolk	1:36:54	52:10:36	<i>O. maritima</i>	24

Materials and methods

Sampling sites

Twenty-one populations were sampled from a range of locations throughout Central and Eastern Britain. Plants from each population were identified based on their morphology as *O. repens*, *O. maritima*, *O. spinosa* and *O. intermedia*. The majority of these sites were allopatric, with one clear sympatric site (Harton Down Hill).

Details of the collection sites are given in Table 2.

Genotyping Restharrow

All 411 individuals were genotyped at 10 SSR loci originally isolated from *O. repens*. DNA preparation and PCR were performed as described previously (Kloda et al., 2004). PCR products were visualized using capillary gel electrophoresis on an ABI-3100 DNA sequencer and the data analysed using Genotyper 2.5 (Applied Biosystems, <http://www.appliedbiosystems.com/>). To avoid introducing inconsistencies due to rounding errors, alleles were sorted using a binning macro (Amos et al., 2007).

Microsatellites analysis and testing of the diploid species *O. spinosa/O. intermedia*

For the diploid species, consistency with Hardy–Weinberg (HW) proportions can be tested for by the exact test (Louis and Dempster, 1987; Guo and Thompson, 1992). A HW exact test for heterozygosity deficit was performed for nine populations of *O. spinosa/O. intermedia* at 10 loci using the programme Genepop (Raymond and Rousset, 1995). Following Bonferroni correction of significance thresholds, the adjusted *P*-value for the 5% nominal level was 0.000625 and for the 1% level the *P*-value was 0.000113.

Linkage disequilibrium (LD) was tested for all pairwise combinations of loci for each population of *O. spinosa* and *O. intermedia*. The tests were carried out using the programme FSTAT (Goudet, 1995) and the log-likelihood ratio G-statistic, using 36 000 random assort-

ments, was recalculated based on the randomized data. The adjusted *P*-value for the 5% nominal level was 0.001111 and for the 1% nominal level was 0.000222. No such tests are readily available for tetraploid data, so the tests were limited to the diploid data set.

Spatial genetic structuring among British Restharrow populations

The possibility that Restharrow populations exhibited spatial genetic structuring was investigated through the use of the pairwise genetic distance statistic, *Rho* (ρ). Ronfort et al. (1998) derived ρ for tetraploid populations as an alternative to *Fst*, which is inappropriate for tetraploids. *Rho* is independent of selfing rates, is also unaffected by double reduction rates and was calculated from the microsatellite data using the programme SPA-GeDi (Hardy and Vekemans, 1999, 2002; Hardy et al., 2000). Correlations between the resulting matrices were tested with a Mantel test (Mantel, 1967; Sokal, 1979) using 10 000 permutations in the programme TFPGA (Miller, 1997).

Comparisons of diversity across populations using PCO analysis

A matrix was constructed counting each individual plant sample as a single case and each microsatellite allele as a variable, scored as present (1) or absent (0). The composite genotypes were presented as a single row. A matrix of allele frequencies was also constructed, representing the frequency of each allele for each population. The data sets were explored using Principal Coordinates Analysis (PCO; Gower, 1966), implemented through the Multivariate Statistical Package (Kovach, 2006). This ordination method makes no assumptions about the distribution of the variates or about their population genetics. Euclidian distance was chosen in preference to other distance measures, as it does not class common absence of an allele as a shared characteristic, and was therefore judged to be most appropriate in the

present study, which included highly polymorphic microsatellite data spanning two ploidy levels.

Trees were generated based on Nei's genetic distance (Nei, 1972) using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) in POPGENE 1.32 (Yeh *et al.*, 1999) from a matrix of presence and absence of each allele. As this programme treats '0' as an allelic state and therefore shared absence of an allele as a common characteristic, microsatellite alleles were scored as present (1) or as missing data (-). Trees were visualized using the programme TREEVIEW version 1.6.6 (Page, 1996).

Harton down hill—a sympatric site

The majority of populations sampled were allopatric, potentially as a result of different habit and soil preference (Table 1). The best example of a sympatric site in this study is Harton Down Hill, where populations 12s (*O. spinosa*) and 12r (*O. repens*) were found adjacent to one another on the 'Harton Down Hill' housing estate, separated only by a footpath. The separation was not perfect, with a few *O. spinosa* plants found growing in the *O. repens* patch and *vice versa*. Within the site, the *O. spinosa* (12s) patch covered a greater area than *O. repens* (12r). The *O. repens* and *O. spinosa* plants were classified as showing all the morphological characters typical of the individual groups. The *O. intermedia* (13i) plants were collected from a nearby coastal cliff, approximately 100 m away. Interestingly, the *O. spinosa* patch had set seed heavily, whereas the *O. repens* flowers had withered and fallen off without producing any seeds, appearing to be sterile at the time of sampling for this study. This example is examined in more detail, through PCO analysis, for evidence of gene flow between species where there is an overlap of habitats.

Results

In total, 247 different alleles were detected across the 10 microsatellites genotyped on 411 individuals from Britain. Up to four alleles were found per locus per individual in *O. repens* and *O. maritima*. In the samples of *O. spinosa* and *O. intermedia*, in the vast majority of cases,

no more than two alleles were found per locus per individual. This is consistent with the first two being tetraploid and the last two being primarily diploids.

As microsatellites allow both alleles to be unambiguously identified in a diploid species, tests for deviations from HW equilibrium were carried out using the data from *O. spinosa* and *O. intermedia*. Excess homozygosity was only observed in the population 'Harton Down Hill' at a single locus (locus 3). This population is examined in more detail below, as it is also unusual in other ways. Observed excess homozygosity may be a reflection of microsatellite null alleles or population subdivision. The test was repeated for excess heterozygosity, but no cases were identified. No significant LD was detected between pairs of loci.

The PCO analysis in Figure 1a suggests two broad groups: one includes *O. intermedia* and *O. spinosa* and the other includes *O. repens* and *O. maritima*. There is limited overlap between the two groups and 15% of the variation is explained by the first two axes. No further genetic differentiation was detected within the two groups. The analysis was repeated using the data from each population summarized into allele frequencies (Figure 1b).

Figure 1b shows that, at the population level, *O. spinosa*/*O. intermedia* and *O. repens*/*O. maritima* form two discrete groups in terms of allelic variation with 38.5% of the variation being explained by the first two axes.

An unrooted tree based on Nei's genetic distance is shown in Figure 2. This shows a clear division between taxa, with the exception of *O. spinosa* (12s) and *O. intermedia* (13i), which were placed in the same cluster as samples of *O. repens* (12r). This is in contrast to the PCO analysis of the allele frequency matrix (Figure 1b), where no exceptions were revealed for these populations and might reflect some genetic interaction between these three populations, all of which were found in close proximity at the Harton Down Hill site.

Harton down hill populations—a sympatric site

To investigate this unusual case further, PCO analysis was performed on microsatellite data for individual plants from *O. spinosa*, (12s), *O. repens* (12r) and *O. intermedia* (13i) (Figure 3).

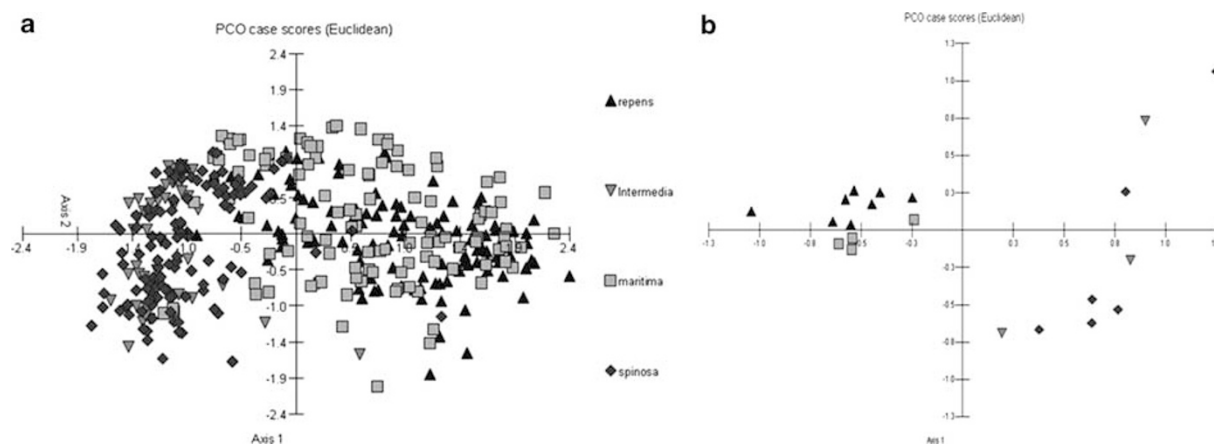


Figure 1 Euclidian PCO analysis for *O. repens*, *O. intermedia*, *O. maritima* and *O. spinosa*. (a) Axes 1 and 2 represent 10.3 and 4.2% of the variation present, respectively, based on an analysis for all individuals. (b) Axes 1 and 2 represent 27.0 and 11.5% of the variation present, respectively, based on population allele frequencies with each data point representing a single population.

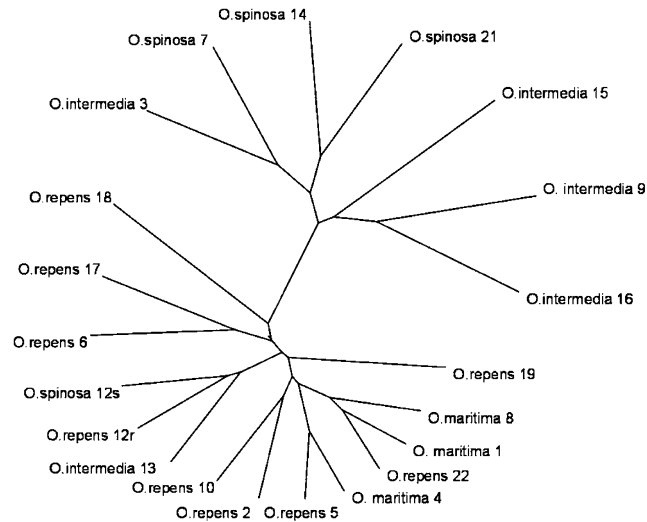


Figure 2 Unrooted UPGMA tree of genetic distance in British Restharrow. The unrooted UPGMA tree was generated in the programme POPGENE from Nei's genetic distances based on microsatellites. UPGMA, Unweighted Pair Group Method with Arithmetic Mean.

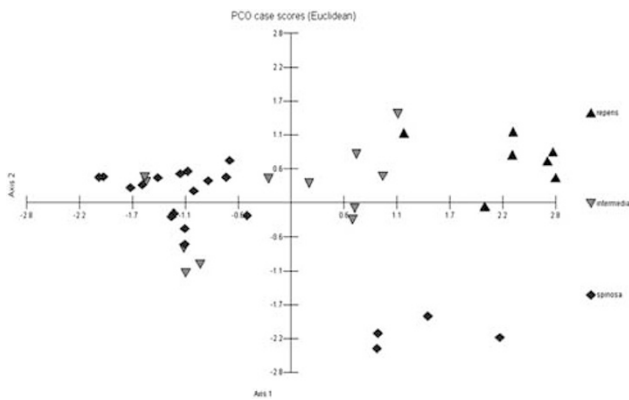


Figure 3 PCO analysis for Harton Down Hill populations. The first two axes of Euclidian PCO analysis for populations *O. repens* (12r), *O. spinosa* (12s) and *O. intermedia* (13i) from a sympatric site. Axes 1 and 2 represent 20.6 and 7.8% of the variation present, respectively.

In the PCO analysis, *O. repens* (12r) formed a relatively tight cluster, while *O. spinosa* (12s) and *O. intermedia* (13i) plants are generally intermixed, despite 12r and 12s being found together, with 13i approximately 100 m away. The PCO also suggests that more genetic variation appears to exist within the *O. spinosa*/*O. intermedia* populations than within the *O. repens* population, although this could possibly be a reflection of relative abundance of plants.

The *O. repens* patch did not appear to be clonal, based on the high level of microsatellite variation found in the samples analysed. An examination of the seed collected from the *O. spinosa* patch confirmed disomic inheritance of markers (data not shown).

Spatial genetic structuring within diploid and tetraploid Restharrow

In accordance with the PCO analysis results, populations were grouped into diploid (*O. spinosa*/*O. intermedia*) and tetraploid (*O. repens*/*O. maritima*) taxa. A Mantels test

showed a highly significant correlation between genetic and spatial distances for tetraploid ($P = 0.0001$) but not diploid species ($P = 0.1095$) using the ρ estimate of pairwise genetic distance (Ronfort *et al.*, 1998; Table 3). Similar results were obtained when the test was repeated using Nei's genetic distance (data not shown).

Discussion

Comparisons across ploidy levels

Jackson (1976) and more recently Ramsey and Schemske (1998) have reviewed the effects of polyploidy on meiosis and genetic transmission for both allopolyploids and autopolyploids. While the situation for pure allopolyploids is usually relatively simple with disomic inheritance, the situation in segmental allopolyploids or autopolyploids is more complex, ranging from mild to complete polysomic inheritance, depending on the degree of chromosomal homology present between the chromosome sets in the polyploid. A number of groups have presented theoretical treatments for dealing with polysomic inheritance in autopolyploids (for example, Glendinning, 1989; Moody *et al.*, 1993; Ronfort *et al.*, 1998; DeSilva *et al.*, 2005) often based around a proposed cytological analysis of bivalent/multivalent ratios in meiosis or on marker inheritance. The development of mapping strategies for autotetraploids has also taken the study of inheritance in such species further (Hackett and Luo, 2003; Luo *et al.*, 2004). However, all methods which aim to estimate missing genotypic data either use defined population structures or make underlying assumptions about species population structure. Kosman and Leonard (2005) recently reviewed the use of similarity coefficients generated from molecular markers and concluded that even microsatellite markers in autopolyploids essentially behaved as dominant markers, making the accurate elucidation of genetic relationships difficult. The comparison of population genetic data within polyploids, let alone across ploidy levels, remains problematic. Here, we have investigated the use

Table 3 Geographical ($\ln x$) and genetic (ρ) distances between populations

Population	1	2	3	4	5	6	7	8	9	10	12r	12s	13	14	15	16	17	18	19	21	22
1		3.17	4.00	3.87	3.87	3.06	4.11	0.12	3.38	1.44	4.12	4.12	4.09	0.86	1.62	1.61	3.59	2.96	4.39	1.88	0.22
2	0.17		2.83	2.90	2.90	0.86	3.11	3.21	1.93	1.74	3.13	3.13	3.12	2.39	1.57	1.57	0.88	0.22	3.13	1.33	3.18
3	0.38	0.39		0.23	0.23	3.64	0.29	3.94	0.92	3.09	0.31	0.31	0.31	3.72	2.95	2.95	2.10	2.79	0.4	2.74	3.99
4	0.19	0.17	0.29		0.00	3.69	0.25	3.81	0.97	3.03	0.27	0.27	0.25	3.63	2.90	2.89	2.21	2.85	0.54	2.70	3.86
5	0.22	0.20	0.29	0.10		3.69	0.25	3.81	0.97	3.03	0.27	0.27	0.25	3.63	2.90	2.89	2.21	2.85	0.54	2.70	3.86
6	0.27	0.14	0.41	0.23	0.27		3.91	3.14	2.73	1.76	3.93	3.93	3.92	2.21	1.65	1.66	1.73	0.85	3.96	1.54	3.08
7	0.33	0.38	0.23	0.26	0.32	0.42		4.05	1.19	3.28	0.02	0.02	0.03	3.88	3.15	3.15	2.39	3.07	0.36	2.95	4.10
8	0.11	0.19	0.39	0.23	0.24	0.25	0.38		3.35	1.48	4.06	4.06	4.03	0.94	1.65	1.64	3.60	3.00	4.33	1.91	0.11
9	0.36	0.37	0.32	0.30	0.32	0.41	0.26	0.40		2.29	1.20	1.20	1.19	2.97	2.13	2.13	1.28	1.88	1.26	1.89	3.38
10	0.14	0.13	0.36	0.21	0.19	0.27	0.34	0.20	0.30		3.29	3.29	3.27	0.73	0.19	0.18	2.19	1.52	3.48	0.47	1.45
12r	0.27	0.22	0.39	0.24	0.26	0.32	0.32	0.34	0.34	0.18		0.00	0.02	3.90	3.16	3.16	2.41	3.09	0.36	2.97	4.11
12s	0.27	0.29	0.23	0.23	0.27	0.35	0.32	0.32	0.18	0.23	0.25		0.02	3.90	3.16	3.16	2.41	3.09	0.36	2.97	4.11
13	0.23	0.27	0.32	0.23	0.29	0.35	0.18	0.32	0.25	0.22	0.26	0.09		3.87	3.14	3.14	2.40	3.08	0.39	2.95	4.09
14	0.39	0.37	0.29	0.33	0.35	0.42	0.33	0.40	0.23	0.33	0.37	0.28	0.36		0.91	0.91	2.91	2.17	4.11	1.19	0.87
15	0.45	0.45	0.31	0.39	0.38	0.48	0.30	0.48	0.07	0.37	0.42	0.21	0.30	0.28		0.01	2.00	1.35	3.34	0.28	1.62
16	0.35	0.35	0.31	0.34	0.35	0.41	0.30	0.37	0.28	0.30	0.35	0.20	0.23	0.30	0.29		2.01	1.36	3.34	0.28	1.62
17	0.20	0.21	0.32	0.17	0.21	0.30	0.31	0.26	0.25	0.11	0.26	0.20	0.19	0.34	0.32	0.32		0.96	2.35	1.72	3.59
18	0.21	0.18	0.48	0.29	0.30	0.34	0.45	0.31	0.44	0.19	0.35	0.36	0.34	0.47	0.56	0.42	0.26		3.11	1.12	2.97
19	0.19	0.22	0.38	0.17	0.24	0.31	0.33	0.28	0.30	0.15	0.17	0.26	0.25	0.38	0.45	0.38	0.16	0.27		3.12	4.38
21	0.31	0.31	0.29	0.28	0.32	0.38	0.27	0.32	0.32	0.27	0.33	0.15	0.22	0.27	0.39	0.27	0.31	0.42	0.32		1.89
22	0.06	0.16	0.38	0.20	0.19	0.26	0.33	0.06	0.36	0.15	0.28	0.26	0.25	0.38	0.45	0.33	0.19	0.24	0.23	0.28	

Geographical distances ($\ln x$) are given above the diagonal and genetic distances (ρ) are given below the diagonal.

of PCO analysis of microsatellite data to provide an alternative approach for such comparisons.

Jackson (1976) summarized a number of potential mechanisms for gene flow between diploid and tetraploid complexes (such as exist for Restharrow) include a triploid bridge (either direction), unreduced diploid pollen ($2n-4n$) or even spontaneous double reduction of pollen ($4n-2n$). The potential of these mechanisms to operate has been supported experimentally (Zohary and Nur, 1959; DeWet, 1971; Vardi, 1974; Tyrl, 1975). However, although technically possible, such mechanisms require a geographical convergence between diploid and tetraploid species—a single case of which we have identified for *O. repens* and *O. spinosa* (Harton Down Hill), but which is unlikely to be common, because of different habit preferences.

Potential gene flow between diploid and tetraploid species has been studied through morphological, isozyme and microsatellites markers (for example, Lumaret and Barrientos, 1990; Gauthier *et al.*, 1998; Raker and Spooner, 2002) generally taking a pragmatic approach using the Nei identity index (Nei, 1972).

Here, we present an approach based on PCO analysis, using data from the Restharrow complex, which includes tetraploids (*O. repens* and *O. maritima*) and diploids (*O. spinosa* and *O. intermedia*). This involved three stages:

(1) The inheritance and reliability of microsatellites derived from the tetraploid species was tested in the diploid species (Kloda *et al.*, 2004). We postulated that the lack of deviations from HW, lack of LD, excess homozygosity (except for one locus in one population) or excess heterozygosity in the heterologous species argued that the markers were likely to exhibit similar performance in the tetraploid species of origin, where formal tests are not possible. That all microsatellite worked equally well in the diploid taxa but were derived from the tetraploid *O. repens* also argues for limited divergence of the two genomes. Deviations from HW could arise due to the Wahlund

effect (Wahlund, 1928), inbreeding, self-fertilization, assortative mating, selection or clonal reproduction.

- (2) We used a PCO analysis to compare relationships based on individual plants within and between *O. spinosa*, *O. intermedia*, *O. repens* and *O. maritima* (Figure 1a).
- (3) We used an allele frequency-based analysis to examine relationships between populations (Figure 1b). We would not expect this to be directly biased by ploidy levels *per se*, as allele frequencies are used in this analysis. However, there will be an under representation of the most common *O. repens/O. maritima* alleles within the allele frequency data, as alleles present in more than one copy in a polyploidy individual will be counted singly. Both individual and population frequency PCO analyses give very similar results, suggesting that the analysis is valid.

The results from this three-step approach suggest that such an analysis is informative, and from this we have concluded that there is little genetic exchange in the UK between the ploidy levels. This simple approach could have general application for such comparisons until tools and techniques to unambiguously distinguish copy numbers of markers in polyploids have been developed and are readily available in plant species. Kosman and Leonard (2005) were unable to identify an ideal approach in this situation and the potential differences identified between PCO and the subsequent UPGMA analysis may be a reflection of this. The PCO does have the advantage that it is identifying the most significant components of the molecular variation, but the disadvantage that it does not utilize all of the available molecular data.

Gene flow in UK Restharrow

Previous authors have suggested several different taxonomic treatments of the common British Restharrow, including the recognition of one, two, three or four species (Table 1). This study provides the first genetic

evidence that there are restrictions to gene flow between the diploid and tetraploid species (*O. spinosa*/*O. intermedia* and *O. repens*/*O. maritima*, respectively), but not between *O. repens* and *O. maritima* nor between *O. spinosa* and *O. intermedia*.

The separate clustering of populations of *O. spinosa*/*O. intermedia* and *O. repens*/*O. maritima* based on microsatellite data suggests they are now largely reproductively isolated. The cause of reproductive isolation is not known, although it might include a number of pre-zygotic or post-zygotic barriers, such as asynchronous flowering, change in pollinator preference or triploid block (Ramsey and Schemske, 1998; Husband and Sabara, 2004).

The significant result of an 'isolation by distance' test for tetraploid (*O. repens*/*O. maritima*) but not for diploid species (*O. spinosa*/*O. intermedia*) would also argue for a lack of interaction between these two groupings, with different factors shaping their population distribution.

Harton down hill—a sympatric site

The possibility of limited gene flow at the Harton Down Hill site (Figures 1b and 2) is interesting, particularly as *O. spinosa* plants set seed the year of sampling and *O. repens* did not. Possible explanations for *O. repens* (12r) plants failing to produce seed in 2003 could be pollinator preference for *O. spinosa*, a specific response to environmental conditions in 2003 or that successful fertilization of *O. repens* (12r) may have been prevented due to an excess of pollination by *O. spinosa*. This is known as minority cytotype exclusion (Levin, 1975) and has been shown to reduce the chances of establishment of tetraploid populations at low frequency in diploid populations, due to fertilization by the predominant haploid pollen on the tetraploid cytotype (Husband, 2000; Baack, 2005). *O. spinosa* is most common at this location, as minority cytotype exclusion would require, but the complete absence of seed in the *O. repens* (12r) plants makes this an unlikely cause for the complete sterility observed. While the species proximity at Harton Down Hill may be an untypical situation (with *O. repens* preferring light, well drained, soils and *O. spinosa* preferring heavy limestone soils) the lack of seed set in the *O. repens* patch, together with seed analysed from *O. spinosa* patch demonstrating disomic inheritance of markers, is strong evidence against extensive genetic interaction between *O. repens* and *O. spinosa*, on the rare occasions when they co-exist in the same geographical location. This supports the conclusions from the PCO analysis.

Whatever the reasons for the reproductive isolation, it is clear that it is not appropriate to consider Restharrow to be a single species in Britain. A more appropriate classification would be into two species, comprising of the diploid species (*O. spinosa*/*O. intermedia*) and tetraploid species (*O. repens*/*O. maritima*). PCO analysis was able to reveal these relationship across ploidy levels and could potentially be used in other species to investigate gene-flow across ploidy levels in plant complexes.

Acknowledgements

We thank Cambio Ltd, who helped support JK on a BBSRC CASE studentship, Dr W Amos, who provided the binning macro for microsatellite analysis, Dr PD Sell

for making his taxonomy of Restharrow available before publication and Dr Francois Balloux for discussions with JK. We thank the Genomics facility at the Department of Genetics, University of Cambridge for working with us to develop the microsatellite data. We acknowledge valuable discussions on gene flow with the late Dr John Barrett. Our thanks to the two anonymous referees for their insightful and helpful comments on the manuscript.

References

- Amos W (1999). A comparative approach to the study of microsatellite evolution. In: Schlotterer C (ed). *Microsatellites: Evolution and Applications*. Oxford University Press: Oxford. pp 66–79.
- Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007). Automated binning of microsatellite alleles: problems and solutions. *Mol Ecol Notes* 7: 10–14.
- Baack EJ (2005). Ecological factors influencing tetraploid establishment in snow buttercups (*Ranunculus adoneus*, Ranunculaceae): minority cytotype exclusion and barriers to triploid formation. *Am J Bot* 92: 1827–1835.
- Cannon P (1981). Systematic studies in the genus *Ononis* L. Leguminosae-Papilionoideae. PhD Thesis, University of Reading.
- Chambers GK, MacAvoy ES (2000). Microsatellites: consensus and controversy. *Comp Biochem Physiol B Biochem Mol Biol* 126: 455–476.
- Crespo MB, Serra L (1993). A new spanish species of the Genus *Ononis* L. Fabaceae. *Bot J Linnean Soc* 111: 37–46.
- DeSilva HN, Hall AJ, Rikkerink E, McNeillage MA, Fraser LG (2005). Estimation of allele frequencies in polyploids under certain patterns of inheritance. *Hered* 95: 327–334.
- DeWet JMJ (1971). Reversible tetraploidy as an evolutionary mechanism. *Evolution* 25: 545–548.
- Ellegren H, Primmer CR, Sheldon BC (1995). Microsatellite evolution: directionality or bias in locus selection. *Nat Genet* 11: 360–362.
- Gauthier P, Lumaret R, Bedecarrats A (1998). Genetic variation and gene flow in Alpine diploid and tetraploid population of Lotus (*L. alpinus* (D.C.) Schleicher/*L. corniculatus* L.) I. Insights from morphological and allozyme markers. *Hered* 80: 683–693.
- Glendinning DR (1989). Some aspects of autotetraploid population-dynamics. *Theor Appl Genet* 78: 233–242.
- Goudet J (1995). FSTAT (Version 1.2): a computer program to calculate F-statistics. *J Hered* 86: 485–486.
- Gower JC (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53: 325–338.
- Greuter W (1986). Proposals to treat names covered under Art69 under Art14. *Taxon* 35: 819–820.
- Guo SW, Thompson EA (1992). Performing the exact test of hardy-weinberg proportion for multiple alleles. *Biometrics* 48: 361–372.
- Gupta M (1980). Anatomy of the flower of *Ononis repens* L. and *Ononis spinosa* L. *J Indian Bot Soc* 59: 366–369.
- Hackett CA, Luo ZW (2003). TetraploidMap: construction of a linkage map in autotetraploid species. *J Hered* 94: 358–359.
- Hardy OJ, Vanderhoeven S, Meerts P, Vekemans X (2000). Spatial autocorrelation of allozyme and quantitative markers within a natural population of *Centaurea jacea* (Asteraceae). *J Evol Biol* 13: 656–667.
- Hardy OJ, Vekemans X (1999). Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Hered* 83: 145–154.
- Hardy OJ, Vekemans X (2002). SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* 2: 618–620.

- Husband BC (2000). Constraints on polyploid evolution: a test of the minority cytotype exclusion principle. *Proc R Soc Lond B* **267**: 217–223.
- Husband BC, Sabara HA (2004). Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae). *New Phytol* **161**: 703–713.
- Ivimey-Cook RB (1968). Investigations into the phenetic relationships between species of *Ononis* L. *Watsonia* **7**: 1–23.
- Jackson RC (1976). Evolution and systematic significance of polyploidy. *Ann Rev Ecol Syst* **7**: 209–234.
- Jarvis C, Ivimey-Cook R, Cannon P (1983). Proposal 696 to reject *Ononis spinosa* Leguminosae. *Taxon* **32**: 314–316.
- Kloda JM, Dean PDG, Macdonald D, Mayes S (2004). Isolation and characterization of microsatellite loci in *Ononis repens*, Leguminosae. *Mol Ecol Notes* **4**: 596–598.
- Kosman K, Leonard KJ (2005). Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Mol Ecol* **14**: 415–424.
- Kovach S (2006). *MVSP—Multivariate Statistical Package v.3.1*. Kovach Computing Services: Anglesey, Wales.
- Langer R, Engler S, Kubelka W (1995). Comparative root anatomy of some perennial taxa of the Genus *Ononis* L. *Pharmazie* **50**: 627–629.
- Levin DA (1975). Minority cytotype exclusion in local plant populations. *Taxon* **24**: 35–43.
- Linnaeus C (1753). *Species plantarum*.
- Louis EJ, Dempster ER (1987). An Exact test for hardy-weinberg and multiple alleles. *Biometrics* **43**: 805–811.
- Lumaret R, Barrientos E (1990). Phylogenetic relationships and gene flow between sympatric diploid and tetraploid plants of *Dactylis glomerata* (Gramineae) P1. *Syst Evol* **169**: 81–96.
- Luo ZW, Zhang RM, Kearsey MJ (2004). Theoretical basis for genetic linkage analysis in autotetraploid species. *Proc Natl Acad Sci* **101**: 7040–7045.
- Mantel N (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res* **27**: 209–220.
- Masterson J (1994). Stomatal size in fossil plants—evidence for polyploidy in majority of Angiosperms. *Science* **264**: 421–424.
- Miller MP (1997). Tools for population genetic analysis (TFPGA) 1.3 a windows program for the analysis of allozyme and molecular population genetic data. Distributed by the author (<http://marks.geneticsoftware.net/tfpga.htm>).
- Moody ME, Mueller LD, Soltis DE (1993). Genetic variation and random drift in autotetraploid populations. *Genet* **134**: 649–657.
- Morisset P (1967). Cytological and taxonomic studies in *Ononis spinosa* L. *O. repens* L and related species. PhD Thesis, University of Cambridge.
- Morisset P (1978). Chromosome numbers in *Ononis* L. species *Vulgares* Sirj. *Watsonia* **12**: 145–153.
- Morton JK (1956). Studies of *Ononis* in Britain: 1. Hybridity in Durham coast colonies of *Ononis*. *Watsonia* **3**: 307–316.
- Nei M (1972). Genetic distance between populations. *Amer Naturalist* **106**: 283–292.
- Page RDM (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput Applicat Biosci* **12**: 357–358.
- Raker CM, Spooner DM (2002). Chilean tetraploid cultivated potato, *Solanum tuberosum*, is distinct from the andean populations: microsatellite data. *Crop Sci* **42**: 1451–1458.
- Ramsey J, Schemske DW (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Ann Rev Ecol and Syst* **29**: 467–501.
- Raymond M, Rousset F (1995). Genepop (version 1.2) population genetics software for exact tests and ecumenicism. *J Hered* **86**: 248–249.
- Robertson A, Newton AC, Ennos RA (2004). Multiple hybrid origins, genetic diversity and population genetic structure of two endemic *Sorbus* taxa on the Isle of Arran, Scotland. *Mol Ecol* **13**: 123–134.
- Rodriguez-Riano T, Ortega-Olivencia A, Devesa JA (1999). Types of androecium in the Fabaceae of SW Europe. *Ann Bot* **83**: 109–116.
- Ronfort JL, Jenczewski E, Bataillon T, Rousset F (1998). Analysis of population structure in autotetraploid species. *Genet* **150**: 921–930.
- Rouy G, Foucaud J (1897). *Flore de France ou description de plantes qui croissent spontanément en France, en Corse et en Alsace-Lorraine*. Derolle: Paris.
- Rowan MG, Dean PD (1972). a-Onocerin and sterol content of 12 species of *Ononis*. *Phytochemistry* **11**: 3263–3265.
- Saint-Martin M (1992). Phytodermology and testa topography in *Ononis* (Papilionaceae). *Willdenowia* **22**: 37–47.
- Sañudo A, Ruiz-Rejón M, Pretel A (1979). Variabilité chromosomique chez les especes d'*Ononis* de la flore espagnole. Note préliminaire. *Webbia* **34**: 535–542.
- Schlötterer C (1998). Microsatellites. In: Hoelzel AR (ed). *Molecular Genetic Analysis of Populations: A Practical Approach*. IRL Press: Oxford. pp 237–261.
- Sell PD, Murrell G (2006). *Flora of Great Britain, Ireland, Isle of Man and the Channel Islands*. Cambridge University Press: Cambridge.
- Sirjaev G (1932). Genus *Ononis* revisio critica. *Beihefte zum Botanische Centralblatt* **49**: 381–665.
- Sokal R (1979). Testing significance of geographic variation patterns. *Syst Zoology* **28**: 227–232.
- Stace C (1997). *New Flora of the British Isles*, 2nd edn. Cambridge University Press: Cambridge, UK.
- Stephens CE (1978). Variation in terminal leaflet shape of *Ononis repens* L. in the British Isles. *Watsonia* **12**: 165–166.
- Stephens CE (1979a). Variation in flower and fruit ratios of *Ononis repens* L. in the British Isles. *Watsonia* **12**: 260.
- Stephens CE (1979b). Some morphological variation in *Ononis spinosa* L. in the British Isles. *Watsonia* **12**: 260–261.
- Tyrl RJ (1975). Origin and distribution of polyploid *Achillea* (Compositae) in western North America. *Brittonia* **27**: 187–296.
- Vardi A (1974). Introgression from tetraploid durum wheat to diploid *Aegilops longissima* and *A. speltoides*. *Hered* **32**: 171–181.
- Wahlund S (1928). The combination of populations and the appearance of correlation examined from the standpoint of the study of heredity. *Hered* **11**: 65–106.
- Wigginton MJ (1999). *British Red Data Books*, 3rd edn. Joint Nature Conservation Committee: Peterborough.
- Wright S (1943). Isolation by distance. *Genet* **28**: 114–138.
- Wright S (1946). Isolation by distance under diverse systems of mating. *Genet* **31**: 39–59.
- Wright S (1951). The genetical structure of populations. *Ann Eugen* **15**: 323–354.
- Yeh FC, Boyle T, Yang RC, Ye ZH, Mao X (1999). *POPGENE*. University of Alberta: Edmonton, Canada.
- Zohary D, Nur U (1959). Natural tetraploids in the Orchard grass *Dactylis glomerata* L. polyploidy complex and their significance for gene flow from diploid to tetraploid levels. *Evolution* **13**: 311–317.