Heredity (2007) 99, 531–544
© 2007 Nature Publishing Group All rights reserved 0018-067X/07 $30.00
www.nature.com/hdy

ORIGINAL ARTICLE

Extremely high cytoplasmic diversity in natural and breeding populations of *Lolium* (Poaceae)

S McGrath[1,2], TR Hodkinson[2] and S Barth[1]
[1]*Teagasc Crops Research Centre, Oak Park, Carlow, Ireland and* [2]*Department of Botany, School of Natural Sciences, Trinity College Dublin, University of Dublin, Dublin, Ireland*

Ten chloroplast microsatellite markers were used to characterise chloroplast genetic diversity at allelic and haplotypic level in 104 accessions of *Lolium perenne*, other *Lolium* species, *Festuca* species and × *Festulolium* cultivars. Furthermore, genetic relationships between the accessions and biogeographic distribution of haplotypes were investigated using a range of Nei's population genetic diversity measures and analysis of molecular variance (AMOVA). An extremely high number (511) of haplotypes was detected in 1575 individuals. Nei's gene diversity values among *L. perenne* accessions ranged between 0 and 0.333. Much of the *L. perenne* European ecotype diversity (61%), as calculated using AMOVA, could be attributed to within-population variance and this is likely caused by, and maintained by, high levels of natural and anthropogenic seed dispersal. Plastid gene pools and maternal lineages for *L. perenne* could be clearly identified. Evidence was found, using AMOVA, to show a likely migration route of *L. perenne* from Southern regions of Europe northwards.
Heredity (2007) **99,** 531–544; doi:10.1038/sj.hdy.6801030; published online 11 July 2007

Keywords: chloroplast; cytoplasm; *Lolium*; microsatellites; plastid; Poaceae

Introduction

Lolium species are highly important forage and turf grasses in Europe and other temperate zones of the world. Perennial ryegrass, *Lolium perenne*, is the most economically important species within *Lolium* and is the main constituent of productive pastures (Humphreys *et al*., 2006). Genetic studies are required for *L. perenne* to better understand the extent of infraspecific variation, how this diversity is partitioned within and among populations, and to help infer phylogeographic patterns. Chloroplast microsatellite (cpSSR) markers have been previously used successfully to assess variation and chloroplast DNA (cpDNA) diversity in a range of other plant species (Provan *et al*., 2001; Flannery *et al*., 2006). cpDNA restriction fragment length polymorphism (RFLP) markers and DNA sequencing have been used to assess the phylogenetic relationships of *L. perenne* to other *Lolium* and *Festuca* species (Darbyshire and Warwick, 1992; Charmet *et al*., 1997; Catalan *et al*., 1997, 2004; Torrecilla *et al*., 2004). These studies showed the separation of narrow-leaved fescues (for example, *Festuca alpine*, *Festuca ovina*) from the broad-leaved fescues (for example, *Festuca arundinacea*, *Festuca pratensis*), with *Lolium* species grouping either close to the broad-leaved fescue group (Catalan *et al*., 1997) or within this group (Darbyshire and Warwick, 1992). Within *Lolium*, self-pollinating species tend to separate phylogenetically from the open pollinating species. In the study undertaken by Catalan *et al*. (2004), two autogamous species, *Lolium canariense* and *Lolium rigidum*, grouped together while the allogamous species *L. perenne* grouped with a second allogamous species, *Lolium multiflorum*.

Genetic characterization of natural and breeding populations of *L. perenne* has so far largely utilized nuclear molecular DNA markers (Cresswell *et al*., 2001; Bolaric *et al*., 2005). Few studies have assessed chloroplast or mitochondrial organelle diversity partly because easily applicable organelle markers have until recently not been easily produced (Huang *et al*., 2002; McGrath *et al*., 2006). Chloroplast genomes of plants have distinct features. They are generally uniparentally inherited, haploid, non-recombinant and have a conserved gene order (Provan *et al*., 2001), making cpDNA a useful tool for studying inter-relationships of plants at many taxonomic levels (Catalan *et al*., 1997; Hodkinson *et al*., 2002). Although cpDNA generally has lower variability than nuclear DNA (Wolfe *et al*., 1987), chloroplast simple sequence repeat (cpSSR) loci have been shown to be polymorphic particularly at mononucleotide repeat loci (Powell *et al*., 1995) and can be used to investigate plant population structure, diversity and differentiation (reviewed by Provan *et al*., 2001). They can be used to monitor the transmission of chloroplast genomes during hybridization and introgression in wild or breeding populations or to characterize plastid genome type for breeding purposes (Flannery *et al*., 2006).

cpSSRs can be used in breeding programmes as cultivar identifiers as an addition to nuclear DNA markers (Joshi *et al*., 1999). Furthermore, detailed

Correspondence: Dr S Barth, Department of Biotechnology, Teagasc Crops Research Centre, Oak Park, Carlow, Ireland.
E-mail: susanne.barth@teagasc.ie
Received 20 December 2006; revised 18 May 2007; accepted 9 June 2007; published online 11 July 2007

characterization of plastid type is essential in studies investigating nucleo-cytoplasmic effects (Hallden *et al*., 1993), since plastid signals controlling nuclear gene expression can have both positive and negative effects on gene expression (Gray, 2005). Several chloroplast genes may have importance for genetic engineering such as those involved in synthesis of fatty acids, amino acids and vitamins (Saski *et al*., 2005) or for the directed manipulation of plant lines in breeding programmes (Daniell *et al*., 2005). Worldwide breeding efforts for the improvement of *L. perenne* and its allies are ongoing and depend upon well-characterized genetic resource collections. Therefore, it is necessary to learn more about the diversity found within it and within closely related species and genera as these form part of the readily accessible gene pool of *L. perenne* for breeding purposes. cpSSR markers have not been used before to study plastid types for plant breeding purposes in *Lolium* species.

L. perenne is thought to have originated in the Near East, with Europe as a secondary centre of origin (Balfourier *et al*., 2000). It has subsequently been introduced to almost all of the rest of the temperate world (Charmet *et al*., 1996). cpDNA RFLP polymorphisms have been used to assess phylogeographic structure in wild-*Lolium* populations and to infer methods and pathways of geographic migration of *Lolium* populations (Balfourier *et al*., 2000). Balfourier *et al*. (2000) recognized three major clusters of haplotypes in their European sample of *Lolium*. Their results suggest a single origin for *Lolium* as well as a geographical structure following an east/west cline, matching known historical processes such as the emergence of agriculture and cereal crops from the Fertile Crescent 10 000 years ago and the spread of these crops towards Europe. As yet, cpDNA SSRs have not been used to study plastid diversity in populations of *L. perenne* or other *Lolium* species; neither have they been used to study the phylogeography of these species or *Festuca* species. Given the agronomic importance of *L. perenne* for European agriculture, migration routes from its centres of origin require investigation.

The aims of this study are to (1) describe cpDNA allelic and haplotypic diversity in natural and breeding populations of *Lolium* including Irish and other European *L. perenne* ecotypes and bred *L. perenne* and $\times$*Festulolium* cultivars; (2) assess the potential of a set of previously developed cpSSR markers (McGrath *et al*., 2006) for plastid genome identification and to assess their value for the study of hybridization and introgression and for the definition of cytoplasmic pools in plant breeding; and (3) determine the level of biogeographic population genetic structure in Irish and European *L. perenne* populations and to gain insights into their phylogeography.

Materials and methods

In total, 104 grass accessions including 78 *L. perenne* accessions were studied (Table 1). These 78 accessions consisted of 30 Irish ecotypes, 32 European ecotypes and 16 cultivars. In addition, 11 other *Lolium* species, 6 *Festuca* species and 9 $\times$*Festulolium* cultivars were used. The term ecotype is used broadly in this article to define locally adapted populations. The Irish ecotypes were selected from the Teagasc Oak Park collection holding 419 *L. perenne* accessions collected from old Irish pasture ecosystems (Connolly, 2000). This collection was made between the years 1980 and 1982 as part of the *Lolium* Core Collection Project that was coordinated by the European Co-operative Programme for Genetic Resources (ECPGR). The populations originated from collection sites where, according to the farmer, no reseeding had been carried out for 50 years or more. For this study, accessions were selected from the Teagasc Oak Park collection to cover a widespread of diverse geographic regions from the Republic of Ireland. The other European accessions were selected to provide a wide European geographic coverage to allow genetic diversity comparisons to be made with the rest of Europe and to help address possible geographic migration routes of *L. perenne*. The included $\times$*Festulolium* cultivars had a defined maternal lineage. The definition of geographic groups was based on the United Nations definition of geographical subregions (United Nations Publication, 1999).

Seeds were germinated on seed testing paper, seedlings transferred to pots and the plants raised in the greenhouse. The ecotype accessions and cultivars were subsequently planted as spaced plants in the field at Oak Park using 50 individual plants per population. Leaf material from generally 16 individuals of each accession was harvested, freeze-dried and ground with a Retsch bead mill. DNA was extracted using a magnetic bead-based method (MagAttract Plant DNA Core kit, Qiagen, Hilden, Germany) as described previously in McGrath *et al*. (2006). PCR reactions were carried out using the primers developed by, and the thermal cycling parameters used in, McGrath *et al*. (2006). Ten of the 12 described primers were used: TeaCpSSR1, TeaCpSSR2, TeaCpSSR3, TeaCpSSR4, TeaCpSSR5, TeaCpSSR7, TeaCpSSR8, TeaCpSSR10, TeaCpSSR11 and TeaCpSSR12. PCR products were analysed using an ABI 3100 automated DNA sequencer and alleles were sized using GeneMapper V3.7 software (Applied Biosystems, Warrington, UK).

A genetic distance matrix was calculated based on the Nei's genetic distance measure (Nei, 1972), using allele data (characters) without size information. Distances based on a stepwise mutation model, such as the $\delta\mu^2$-distance (Goldstein *et al*., 1995; Flannery *et al*., 2006), were not used because the size variation of alleles could be attributed to both SSR length variation and other types of substitutions (non-SSR indels). From this matrix, a dendrogram showing the similarities between populations was constructed using the unweighted pair group method with arithmetic means (UPGMA) method (Sneath and Sokal, 1973) as implemented in POPGENE (Yeh and Boyle, 1997). Bootstrapping analysis was performed on the UPGMA data with 1000 replicates as implemented in NTSYSpc V2.2 software (Rohlf, 2005).

A geographic distance matrix was constructed, using the 56 accessions where an exact geographical origin was known (Table 1). A Mantel test was used to correlate the pairwise comparisons in the geographic distance matrix and the genetic distance matrix using NTSYSpc V2.2 (Rohlf, 2005). A total of 10 000 permutations were employed to test for significance.

Since polymorphisms in the chloroplast genome are considered to be linked, each haplotype can be con-

Table 1 Name, source, original location and genetic diversity information on accessions of *Lolium*, ×*Festulolium* and *Festuca* used in this study

Species	*Accession number*	*Group*[a]	*Country of origin*	*Location*	*Latitude*	*Longitude*	*Seed source*	N	H	N *Haplotypes/ population*	N *Unique haplotypes*
L. perenne	IRL-OP-02337	I 1	Ireland	Kellistown Farm, Carlow	N 52.47.60	W 06.49.73	Teagasc Oak Park	15	0.180	9	2
L. perenne	IRL-OP-02059	I 2	Ireland	Moyneroe, Scarrif, Clare	N 52.54.35	W 08.30.32	Teagasc Oak Park	16	0.202	12	1
L. perenne	IRL-OP-02007	I 3	Ireland	Bromcloc, Bantry, Cork	N 51.39.95	W 09.31.07	Teagasc Oak Park	16	0.177	11	2
L. perenne	IRL-OP-02011	I 4	Ireland	Crowleys Pub, The Square, Bantry, Cork	N 51.42.15	W 09.27.67	Teagasc Oak Park	14	0.207	11	0
L. perenne	IRL-OP-02015	I 5	Ireland	South Ring, Clonakilty, Cork	N 51.37.10	W 08.53.71	Teagasc Oak Park	16	0.268	15	6
L. perenne	IRL-OP-02048	I 6	Ireland	Carrigeen, Conna, Old Kents, Fermoy, Cork	N 52.05.88	W 08.03.50	Teagasc Oak Park	16	0.202	10	2
L. perenne	IRL-OP-02192	I 7	Ireland	Horse Island, Roaring Water Bay, West Cork	N 51.30.80	W 09.29.03	Teagasc Oak Park	16	0.197	11	1
L. perenne	IRL-OP-02312	I 8	Ireland	Fortlands House, Charleville, Cork	N 52.20.86	W 08.42.27	Teagasc Oak Park	16	0.219	12	3
L. perenne	IRL-OP-02320	I 9	Ireland	Clonakilty, Co. Cork	N 51.37.10	W 08.53.71	Teagasc Oak Park	16	0.202	11	3
L. perenne	IRL-OP-02064	I 10	Ireland	Kilreekill, Loughrea, Galway	N 53.13.23	W 08.28.75	Teagasc Oak Park	16	0.190	10	3
L. perenne	IRL-OP-02078	I 11	Ireland	Ballycahalan, Peterswell, Galway	N 53.05.64	W 08.36.72	Teagasc Oak Park	18	0.324	15	5
L. perenne	IRL-OP-02230	I 12	Ireland	Clough, Cummer, Tuam, Galway	N 53.27.11	W 08.53.29	Teagasc Oak Park	16	0.333	15	8
L. perenne	IRL-OP-02128	I 13	Ireland	Mahera Beg, Commonage North, Castlegregory, Kerry	N 52.17.79	W 10.01.38	Teagasc Oak Park	16	0.122	5	3
L. perenne	IRL-OP-02538	I 14	Ireland	Colt, Ballyroan, Laois	N 52.58.08	W 07.20.70	Teagasc Oak Park	16	0.198	8	2
L. perenne	IRL-OP-02274	I 15	Ireland	Buffanoka, Cappamore, Limerick	N 52.39.29	W 08.18.62	Teagasc Oak Park	16	0.174	9	1
L. perenne	IRL-OP-02480	I 16	Ireland	Inch St, Lawrence, Caherconlish, Limerick	N 52.35.47	W 08.30.99	Teagasc Oak Park	16	0.276	16	5
L. perenne	IRL-OP-02442	I 17	Ireland	Doughmakean, Roonagh, Westport, Mayo	N 53.44.74	W 09.53.69	Teagasc Oak Park	16	0.270	15	3
L. perenne	IRL-OP-02444	I 18	Ireland	Barnabawn, Killadoon, Westport, Mayo	N 53.41.48	W 09.54.91	Teagasc Oak Park	16	0.191	10	1
L. perenne	IRL-OP-02068	I 19	Ireland	Ballycommon, Tullamore, Offaly	N 53.17.50	W 07.23.11	Teagasc Oak Park	15	0.287	14	7
L. perenne	IRL-OP-02241	I 20	Ireland	Clonohill, Birr, Offaly	N 53.05.20	W 07.53.73	Teagasc Oak Park	16	0.253	13	6
L. perenne	IRL-OP-02419	I 21	Ireland	Johnstown,Cornafulla, Athlone, Roscommon	N 53.38.05	W 09.30.71	Teagasc Oak Park	16	0.253	14	4
L. perenne	IRL-OP-02258	I 22	Ireland	The Lawn, Drum, Tipperary	N 52.46.32	W 07.52.89	Teagasc Oak Park	16	0.191	11	3
L. perenne	IRL-OP-02272	I 23	Ireland	Ballycrana, Kilross, Tipperary	N 52.25.28	W 08.15.88	Teagasc Oak Park	16	0.301	16	6
L. perenne	IRL-OP-02250	I 24	Ireland	Glown, Upperchurch, Tipperary	N 52.42.55	W 08.07.99	Teagasc Oak Park	17	0.260	13	3
L. perenne	IRL-OP-02267	I 25	Ireland	Ballyhoulihan, Emly, Tipperary	N 52.26.88	W 08.22.06	Teagasc Oak Park	16	0.292	15	4
L. perenne	IRL-OP-02269	I 26	Ireland	Ballycurrane, Emly, Tipperary	N 52.27.42	W 08.22.95	Teagasc Oak Park	17	0.248	15	2
L. perenne	IRL-OP-02173	I 27	Ireland	Deerpark, Lismore, Waterford	N 52.08.04	W 07.55.62	Teagasc Oak Park	16	0.273	12	3
L. perenne	IRL-OP-02483	I 28	Ireland	Edwardstown, Cleriestown, Wexford	N 52.16.20	W 06.38.25	Teagasc Oak Park	16	0.262	16	1

Table 1 Continued

Species	*Accession number*	*Group*[a]	*Country of origin*	*Location*	*Latitude*	*Longitude*	*Seed source*	N	H	N *Haplotypes/ population*	N *Unique haplotypes*
L. perenne	IRL-OP-02491	I 29	Ireland	Heath Park, Newbawn, Wexford	N 52.23.32	W 06.48.61	Teagasc Oak Park	16	0.248	14	2
L. perenne	IRL-OP-02018	I 30	Ireland	Ballynure Demesne, Grangecon, Wicklow	N 52.59.95	W 06.44.92	Teagasc Oak Park	16	0.286	14	6
L. perenne	GR 5092	Δ1	Germany	Malchow/Poel	N 54.00.00	E 11.28.00	IPK Gatersleben	16	0.199	12	2
L. perenne	PI 598445	Δ2	Netherlands	Unknown	N 53.07.00	E 07.02.00	GRIN	12	0.146	8	0
L. perenne	ABY-Ba 12896	Δ3	Denmark	Unknown	N 55.00.00	E 09.46.59	IGER	16	0.128	8	5
L. perenne	NGB14250	Δ4	Sweden	Unknown	N 57.45.50	E 14.51.25	Nordic Gene Bank	12	0.285	11	3
L. perenne	16-7-62-2 Nordic	Δ5	Norway	Sola	N 58.54.00	W 05.34.99	Teagasc Oak Park	16	0.268	13	6
L. perenne	PI 619024	Δ6	England	Unknown	N 53.17.00	W 01.46.00	GRIN	12	0.106	6	2
L. perenne	W6 9339	Δ7	Wales	Unknown	N 51.57.00	W 03.03.00	GRIN	16	0.247	14	2
L. perenne	PI 610958	□8	Tunisia	Unknown	N 36.53.42	E 09.11.13	GRIN	16	0.260	14	7
L. perenne	ABY-Ba 11315	□9	Morocco	Unknown	N 31.30.00	W 09.48.00	IGER	16	0.206	12	5
L. perenne	E1	□10	Egypt	Unknown	Unknown	Unknown	PGG-Wrightson	16	0.066	4	3
L. perenne	W6 11325	▲11	Turkey	Karabuk, Ankara	N 41.07.12	E 32.22.12	GRIN	16	0.114	6	1
L. perenne	PI 598512	▲12	Turkey	Antalya	N 36.54.45	E 30.41.23	GRIN	16	0.188	9	2
L. perenne	PI 547390	▲13	Iran	Karaj	N 35.28.48	E 51.00.00	GRIN	12	0.118	5	1
L. perenne	PI 317452	▲14	Afghanistan	North of Hari Rud River, 4 miles West of Besha	N 34.46.00	E 63.46.00	GRIN	16	0.165	9	1
L. perenne	No 10 Spain	■15	Spain	Unknown	Unknown	Unknown	Teagasc Oak Park	16	0.250	13	6
L. perenne	3408 Italy	■16	Italy	Unknown	Unknown	Unknown	Teagasc Oak Park	16	0.264	15	4
L. perenne	W6 16127	■17	Italy	Sardinia	N 40.34.37	E 09.12.18	GRIN	15	0.191	9	8
L. perenne	3013 Romania	■18	Romania	Unknown	Unknown	Unknown	Teagasc Oak Park	17	0.235	15	3
L. perenne	3199 Romania Podoloni	■19	Romania	Unknown	Unknown	Unknown	Teagasc Oak Park	16	0.235	14	0
L. perenne	920 Bulgaria	■20	Bulgaria	Unknown	Unknown	Unknown	Teagasc Oak Park	16	0.206	11	0
L. perenne	PI 418701	■21	Yugoslavia	Prizren	N 42.13.00	E 22.44.00	GRIN	16	0.120	6	1
L. perenne	ABY-Ba 11478	■22	Greece	Unknown	N 38.00.00	E 22.10.00	IGER	16	0.132	14	3
L. perenne	W6 9286	○23	France	Unknown	N 47.33.00	E 04.28.00	GRIN	16	0.155	9	3
L. perenne	ABY-Ba 11514	○24	France	Unknown	N 49.57.00	E 02.46.00	IGER	16	0.182	10	3
L. perenne	CPI 44924	○25	France	Arles	N 43.40.01	E 04.37.58	PGG-Wrightson	16	0.205	11	7
L. perenne	GR 5095	○26	Germany	Kempten	N 47.49.00	E 10.19.59	IPK Gatersleben	16	0.205	12	3
L. perenne	GR 5105	○27	Germany	Kempten	N 47.49.59	E 10.15.00	IPK Gatersleben	16	0.238	12	5
L. perenne	PI 274637	•28	Poland	Lublin	N 51.13.48	E 22.33.00	GRIN	16	0.158	10	1
L. perenne	PI 267058	•29	Poland	Warszawa	N 52.35.00	E 21.05.00	GRIN	16	0.099	5	1
L. perenne	PI 182857	•30	Czech Republic	Central Bohemia	Unknown	Unknown	GRIN	16	0.232	6	6
L. perenne	PI 321397	•31	Czech Republic	Central Bohemia	Unknown	Unknown	GRIN	16	0.232	12	5
L. perenne	IV-51-161 Hungary	•32	Hungary	Unknown	Unknown	Unknown	Teagasc Oak Park	16	0.245	14	6
L. perenne	cv. Aurora	V 1	NA	NA	NA	NA	IGER	17	0.249	12	1
L. perenne	cv. Barlenna	V 2	NA	NA	NA	NA	Barenbrug Holland BV	16	0.229	10	5
L. perenne	cv. Cancan	V 3	NA	NA	NA	NA	DLF-Trifolium	16	0.302	15	6
L. perenne	cv. Cashel	V 4	NA	NA	NA	NA	Teagasc	16	0.172	5	1
L. perenne	cv. Fennema	V 5	NA	NA	NA	NA	Norddeutsche Pflanzenzucht	16	0.207	11	3
L. perenne	cv. Greengold	V 6	NA	NA	NA	NA	Teagasc	17	0.161	9	0
L. perenne	cv. Magician	V 7	NA	NA	NA	NA	Teagasc	16	0.254	14	3

Table 1 Continued

Species	*Accession number*	*Group*[a]	*Country of origin*	*Location*	*Latitude*	*Longitude*	*Seed source*	N	H	N *Haplotypes/ population*	N *Unique haplotypes*
L. perenne	cv. Millenium	V 8	NA	NA	NA	NA	Teagasc	16	0.157	9	0
L. perenne	cv. Navan	V 9	NA	NA	NA	NA	DARDNI	16	0.272	11	4
L. perenne	cv. Odenwaelder	V 10	NA	NA	NA	NA	IPK Gatersleben	16	0.259	14	4
L. perenne	cv. Portstewart	V 11	NA	NA	NA	NA	DARDNI	16	0.193	9	3
L. perenne	cv. Premo	V 12	NA	NA	NA	NA	Mommersteeg International BV	16	0.167	9	0
L. perenne	cv. S24	V 13	NA	NA	NA	NA	IGER	17	0.199	12	5
L. perenne	cv. Sarsfield	V 14	NA	NA	NA	NA	Teagasc	16	0.238	12	2
L. perenne	cv. Shandon	V 15	NA	NA	NA	NA	Teagasc	16	0.213	11	2
L. perenne	cv. Talbot	V 16	NA	NA	NA	NA	Van der Have Grasses BV	16	0.235	13	3
L. canariense	PI 320544	L 1	Canary Islands	Iqueste de San Andres, Tenerife	N 110.36.00	E 12.00.29	GRIN	16	0.239	7	7
L. hybridum	ABY-Ba 13122	L 2	Portugal	Unknown	N 40.56.00	W. 07.33.00	IGER	16	0.157	9	9
L. hybridum	GR11849/94	L 3	NA	Unknown	Unknown	Unknown	IPK Gatersleben	8	0.109	5	3
L. multiflorum	GR11855/98	L 4	NA	Unknown	Unknown	Unknown	IPK Gatersleben	8	0.231	6	6
L. persicum	PI 229764	L 5	Iran	Unknown	Unknown	Unknown	GRIN	16	0.065	3	4
L. remotum	GR11839/99a	L 6	Germany	Unknown	Unknown	Unknown	IPK Gatersleben	8	0.052	3	3
L. rigidum	GR11848/91	L 7	Iran	Between Kalafabad and Ramhormos	Unknown	Unknown	IPK Gatersleben	8	0.066	4	4
L. subulatum	PI 197310	L 8	Argentina	Unknown	Unknown	Unknown	GRIN	16	0.156	6	7
L. temulentum	ABY-Ba 13643	L 9	Morocco	Unknown	N 35..34.00	W 05.22.00	IGER	16	0.227	13	11
L. temulentum	ABY-Ba 8917	L 10	Iran	Unknown	N 52.19.00	E 36.25.59	IGER	16	0.174	9	9
L. temulentum	GR11880/82	L 11	Italy	South of Monte Mutry, 1 km east of Santa Crocella	Unknown	Unknown	IPK Gatersleben	8	0.147	5	6
× *Festulolium braunii*	cv. Perun	F 1	NA	NA	NA	NA	Plant Breeding Station Hladke Zivotice	16	0.240	13	9
× *Festulolium braunii*	cv. HD 14 DK	F 2	NA	NA	NA	NA	Plant Breeding Station Hladke Zivotice	16	0.247	13	6
× *Festulolium braunii*	cv. Paulita	F 3	NA	NA	NA	NA	Plant Breeding Station Hladke Zivotice	16	0.207	11	4
× *Festulolium braunii*	cv. Achilles	F 4	NA	NA	NA	NA	Plant Breeding Station Hladke Zivotice	16	0.187	10	1
× *Festulolium Lolium multiflorum* × *Festuca arundinacea BC to F. arundinacea*	cv. Lesana	F 5	NA	NA	NA	NA	Plant Breeding Station Hladke Zivotice	16	0.213	13	0
× *Festulolium Lolium multiflorum* × *Festuca arundinacea BC to F. arundinacea*	cv. Korina	F 8	NA	NA	NA	NA	Plant Breeding Station Hladke Zivotice	16	0.215	11	0
× *Festulolium Lolium multiflorum* × *Festuca arundinacea BC to F. arundinacea*	cv. Felina	F 9	NA	NA	NA	NA	Plant Breeding Station Hladke Zivotice	12	0.214	8	1
× *Festulolium Lolium multiflorum* × *Festuca arundinacea BC to L. multiflorum*	cv. Becva	F 6	NA	NA	NA	NA	Plant Breeding Station Hladke Zivotice	16	0.251	13	1

Table 1 Continued

Species	*Accession number*	*Group*[a]	*Country of origin*	*Location*	*Latitude*	*Longitude*	*Seed source*	N	H	N *Haplotypes/ population*	N *Unique haplotypes*
× *Festulolium Lolium multiflorum* × *Festuca arundinacea BC to L. multiflorum*	cv. Lofa	F 7	NA	NA	NA	NA	Plant Breeding Station Hladke Zivotice	16	0.273	15	10
Festuca arundinacea	cv. Dovey	NL 1	NA	NA	NA	NA	Barenbrug Holland BV	8	0	1	1
Festuca gigantean	PI 440362	NL 2	Kazakhstan	Near river between Alma Ata and and Medeo	Unknown	Unknown	GRIN	8	0.122	5	5
Festuca ovina	PI 634304	NL 3	China	Xinjiang	N 43.28.05	E 81.06.39	GRIN	16	0.148	9	7
Festuca pratensis	cv. Northland	NL 4	NA	NA	NA	NA	PGG-Wrightson	8	0	1	1
Festuca rubra	IRL-OP-02174	NL 5	Ireland	Kilworth Range, Fermoy, Cork	N 52.12.88	W 08.15.80	Teagasc Oak Park	8	0.115	4	3
Festuca vivipara	PI 251118	NL 6	Yugoslavia	Unknown	Unknown	Unknown	GRIN	16	0.109	5	3

Abbreviations: BC, Backcross; *H*, Nei's (1973) gene diversity; L, *Lolium* species; *N*, number of individuals; NA, not applicable; NL, non-*Lolium* species; V, *Lolium perenne* variety.
[a]Group: I, Irish ecotype; Δ, Northern Europe group 1; □, North Africa group 2; ▲, Near East group 3; ■, Southern Europe group 4; ○, Western Europe group 5; λ, Eastern Europe group 6.

sidered as one allele at a single haploid locus. Allele size information was used to construct haplotypes. Haplotype frequencies were used to calculate Nei's (1973) gene diversity (H) and the total diversity (H_t), the mean within-population diversity (H_s) and the coefficient of variation (G_{st}) following the procedures of Nei (1987) using the program POPGENE (Yeh and Boyle, 1997).

An analysis of molecular variance (AMOVA; Excoffier *et al*., 1992) was carried out with ARLEQUIN 2.0 software (Schneider *et al*., 2000) based on the number of different haplotypes, which is the equivalent of a weighted F_{st} over all loci when estimating genetic structure (Weir and Cockerham, 1984; Michalakis and Excoffier, 1996). The level of significance for variance component estimates was calculated by non-parametric permutation procedures using 1000 permutations. The data were partitioned in several combinations to display among- and within-population variation of Irish and European *L. perenne* accessions, to assess biogeographic differentiation and to address possible migration routes of *L. perenne*.

AMOVAs were calculated to test for evidence of geographic patterns of genetic structuring. In addition, we used AMOVAs to test if there was evidence of migration of *Lolium* (a) following a Mediterranean route, this involved comparisons of Near Eastern vs Southern European and Southern European vs Western European *L. perenne* accession groups; (b) following a Danubian migration route, this involved comparisons of Near Eastern vs Southern European, Southern European vs Eastern European and Eastern European vs Northern European *L. perenne* accession groups; (c) following a Northern African route, this involved comparisons of Near Eastern vs Northern African and Northern Africa vs Southern European *L. perenne* accession groups; (d) following a northerly retreat route of the glaciers after the last ice age, this involved comparison of all *L. perenne* ecotype groups south of the Alps (Near East, North Africa and Southern Europe) against all *L. perenne* ecotype groups North of the Alps (Northern Europe, Eastern Europe, Western Europe, Ireland); and (e) following routes into Ireland consistent with migration from neighbouring geographical regions, this involved comparisons of Irish *L. perenne* ecotypes with three European accession groups (Southern Europe, Western Europe and Northern Europe).

An Edward's Venn diagram was constructed to display shared haplotypes between six accession groups: Irish *L. perenne* ecotypes, European/Near Eastern *L. perenne* ecotypes, *L. perenne* cultivars, other *Lolium* species, × *Festulolium* cultivars and *Festuca* species. A second Edward's Venn diagram was constructed to display shared haplotypes between seven *L. perenne* accession groups: Irish ecotypes, Northern European ecotypes, Western European ecotypes, Southern European ecotypes, Eastern European ecotypes, North African ecotypes and Near Eastern ecotypes.

From the genetic distance matrix, a dendrogram showing the similarities between six major groups of accessions was constructed using UPGMA (Sneath and Sokal, 1973) as implemented in POPGENE (Yeh and Boyle, 1997). Bootstrapping analysis was performed on the UPGMA data with 1000 replicates as implemented in NTSYSpc V2.2 software (Rohlf, 2005).

Results

Allelic variation

All ten cpSSR marker loci used in this analysis of 1575 individuals were found to be polymorphic, ranging from marker TeaCpSSR7 with 4 alleles to marker TeaCpSSR3 with 22 alleles (Table 2). The distribution of alleles in the populations varied between *L. perenne* ecotypes, *L. perenne* cultivars and the groups of other species. At locus TeaCpSSR8, there was only one allele present in the × *Festulolium* cultivars, but ten alleles present in Irish *L. perenne* ecotypes. Loci TeaCpSSR3 and TeaCpSSR8 had the largest number of alleles for *L. perenne* ecotypes. Loci TeaCpSSR2, TeaCpSSR3 and TeaCpSSR5 had the largest number of alleles for *L. perenne* cultivars. Locus TeaCpSSR3 was extremely rich in alleles, including 11 alleles for the other tested *Lolium* species.

Generally, there were more alleles unique to *L. perenne* ecotypes than in all the other species groups. Marker locus TeaCpSSR8 was the richest locus for unique alleles in *L. perenne* in general. There were five alleles unique to the Irish *L. perenne* ecotypes, five alleles unique to the European/Near Eastern ecotypes and one allele unique to the other *Lolium* species. None of the alleles at this locus was unique to the × *Festulolium* cultivars or to *Festuca* species. Locus TeaCpSSR7 was an exception with no allele being unique to *L. perenne* accessions. None of the alleles, across all loci, was diagnostic by itself for a single population, but some were for a defined group of populations. However, unique alleles were present only in groups containing more than one allele. Three alleles were unique for non-*L. perenne Lolium* species (at loci TeaCpSSR3 and TeaCpSSR8). Generally across all ten marker loci, no unique alleles for *Festuca* and × *Festulolium* accessions were found (Table 2).

An UPGMA dendrogram showing the similarities between all 104 populations was constructed and it produced groupings that were similar to the UPGMA analysis of 11 defined groups of accessions (Figure 1). For this reason only the latter dendrogram is shown. The dendrogram of all 104 accessions is provided as Supplementary Information (Supplementary Material A).

The UPGMA dendrogram showing the similarities between the 11 groups of accessions was constructed to support the AMOVA analysis (below) and to investigate the broad-scale geographical structuring (Figure 1). The group of *Festuca* species were outlying all other groups (Figure 1, group I). The rest of the dendrogram was split into two major groups (Figure 1, groups II and III). The first group (II) contained the Irish *L. perenne* ecotypes, the *L. perenne* cultivars, the *Lolium* species and the × *Festulolium* cultivars. The second group (III) contained all European/Near Eastern *L. perenne* ecotypes and could be split into two subgroups (IIIa and IIIb). Subgroup IIIa consisted of the Southern European, Western European and Northern European ecotypes, while subgroup IIIb consisted of the North African, Near Eastern and Eastern European ecotypes. There was moderate-to-good bootstrap support for many branches of the tree.

A Mantel test to check for a correlation between genetic distances and geographic distances among 56 accessions out of 62 ecotypes resulted only in a weak and nonsignificant correlation ($r = 0.338$).

Haplotypic variation

The 104 tested populations had a large amount of haplotypic variation (Table 1 and Figure 2). Of the 511 haplotypes present, 366 of these were unique to individual populations. Generally with a few exceptions, each of the 104 populations had a range of unique haplotypes (Table 1). Nine populations had no unique haplotypes, while four populations (*L. temulentum* L10, *F. arundinacea*, *F. gigantea* and *F. pratensis*) were composed of completely unique haplotypes (Table 1). No single haplotype was present in all groups of populations. One hundred and twelve of the haplotypes were only present in Irish *L. perenne* ecotypes (Figure 2) of which 98 were unique to single populations. Thus 15 haplotypes were shared only among Irish *L. perenne* ecotypes and were diagnostic for these accessions (Supplementary Material B). One hundred and twenty-four haplotypes were unique to European *L. perenne* ecotypes of which 105 were unique to single populations. Twenty haplotypes could be considered as diagnostic for European *L. perenne* ecotype accessions (Supplementary Material B), especially for ecotypes of the Northern European and Western European geographical regions. Forty-five haplotypes were found only in *L. perenne* cultivars of which 42 haplotypes were unique to single populations. Three haplotypes were shared among *L. perenne* cultivars (Supplementary material B), and were found in cultivars 'Aurora', 'Cancan', 'Magician' and 'Shandon'. Fifty-two unique haplotypes were detected in × *Festulolium* cultivars of which 32 were unique to single populations. Seventy-four haplotypes were specific to other *Lolium* species of which 69 were unique to single populations. Twenty-one haplotypes were found to be specific for *Festuca* species of which 20 were unique to single populations. In total, 29 haplotypes were shared between × *Festulolium* accessions and *Lolium* accessions, of which 26 haplotypes were shared with *L. perenne* ecotypes (Figure 2).

Nei's gene diversity values (H; Nei, 1972) among *L. perenne* accessions ranged between 0 (NL1 and NL4) and 0.333 with the Irish ecotype I 12 (Table 1). Only one haplotype was present in the *F. arundinacea* and *F. ovina* populations, thus their H values were 0. The lowest H value of 0.052 was found in a *L. remotum* population (L6, Table 1). H values were higher in the Irish and European/Near Eastern *L. perenne* ecotypes ranging between 0.211 and 0.219. In the other groups of accessions, values for cultivars were $H = 0.145$, other *Lolium* species $H = 0.187$, × *Festulolium* cultivars $H = 0.086$ and *Festuca* species $H = 0.148$. The total gene diversity (H_t; Table 3) ranged from Near Eastern *L. perenne* accessions ($H_t = 0.233$) to Southern European *L. perenne* accessions ($H_t = 0.359$). H_t values for other accession groups were within this range. The gene diversity within subdivided populations (H_s) ranged between 0.077 for *Festuca* species and 0.238 for Irish *L. perenne* ecotypes. The highest H_s value in the Irish *L. perenne* ecotypes was closely followed by the value for × *Festulolium* accessions ($H_s = 0.227$). The H_s value for Near Eastern *L. perenne* ecotypes was the lowest ($H_s = 0.146$) among *L. perenne* accessions. As G_{st} becomes closer to one, the populations within the groups are more markedly different from each other. G_{st} values ranged from 0.238 in the Irish *L. perenne* ecotypes to 0.716 in the *Festuca* species.

Table 2 Number of alleles (and unique alleles) per locus in each group of accessions

	N	*TeaCpSSR1*	*TeaCpSSR2*	*TeaCpSSR3*	*TeaCpSSR4*	*TeaCpSSR5*	*TeaCpSSR7*	*TeaCpSSR8*	*TeaCpSSR10*	*TeaCpSSR11*	*TeaCpSSR12*	*Total number of alleles/group*
Irish *Lolium perenne* ecotypes	480	6 (2)	6 (1)	11 (1)	6 (1)	7 (3)	3	10 (5)	4 (1)	3 (1)	4 (2)	60
European and other geographic regional *L. perenne* ecotypes (total)	496	3 (1)	8 (3)	16 (3)	5	3 (1)	2	9 (5)	4 (2)	4 (2)	4 (1)	58
Northern Europe Δ	100	2	2	11	3	2	2	3	1	2	3	31
North Africa □	48	2	3	5	3	2	2	3	2	1	2	25
Near East ▲	60	2	4	4	2	2	2	1	1	2	4	24
Southern Europe ■	128	3	4	8	4	2	2	3	3	2	2	33
Western Europe ○	80	2	5	6	2	2	2	4	2	3	3	31
Eastern Europe •	80	2	3	5	3	3	2	1	1	2	2	24
L. perenne varieties	259	5 (1)	5	8	4 (1)	5 (2)	2	2	3	3 (1)	4 (2)	41
× *Festulolium*	140	2	1	8	2	2	2	1	1	1	2	22
Other *Lolium* species	136	2	3	11 (2)	2	2	4	2 (1)	1	3	3	33
Festuca species	64	1	1	5	2	2	2	4	2	1	1	21
Total number of alleles/locus		8	10	22	7	10	4	17	6	7	8	99

Abbreviation: *N*, number of individuals.
The numbers of unique alleles are shown in parentheses.

AMOVA analysis

AMOVA analysis was carried out on 15 different subgroups of accessions, to test for differences in genetic structure within and between Irish and European ecotypes, to test possible geographic migration routes and to test for differences between and within *L. perenne* accessions and × *Festulolium* cultivars (Table 4). In general, the variation among populations accounted for less of the total variation than that found within populations, for example, for the Irish and European *L. perenne* ecotype comparison, the within-population variation accounted for 63% of total variation. Among-population variation accounted for 26% of total variation and among-group variation accounted for 11% of total variation. These results were comparable to the variation found within and among partitions in the *L. perenne* and × *Festulolium* comparison (64% within-population variation, 17% among-population variation and 19% among-group variation). However, there was more variation among groups and less variation at the among-population level in the *L. perenne* and × *Festulolium* comparison (Table 4).

Within-population variation of Irish *L. perenne* ecotypes accounted for 82% of the variation and among-population variation accounted for 18%. For the European *L. perenne* ecotypes, the situation was different. The within-population variation accounted for only 61% of the total variation, and the among-population variation accounted for 39% (Table 4). For all AMOVA calculations, all results were highly significant ($P \leqslant 0.001$). For analyses comparing the phylogeographic structure in the European/Near Eastern ecotypes, the percentages of variation accounted for by among- and within-population variation were similar for all the partitions tested (data not shown).

For all calculations comparing migration routes among-group and among-population variation was highly significant ($P \leqslant 0.001$). All within-population variations were not significant. In both tests for evidence of a Mediterranean migration route, among-group variation was $\leqslant 0$ (Table 4). For the two tests investigating phylogeographic structure on a possible Danubian migration route (Southern vs Western Europe and Southern vs Eastern Europe), among-group variations were low but significant, 3 and 4%, respectively (Table 4). Similarly, for both tests investigating phylogeographic structure on a possible North African migration route, among-group variations were low but significant, 2 and 4%, respectively (Table 4). A post-glacial migration route appeared to be possible since among-group variation for this partition was 0 ($P \leqslant 0.001$). For migration into Ireland from three neighbouring geographical groups, the lowest among-group variation was found in the partition between Southern European and Irish *L. perenne* ecotypes with only 5%. This result was confirmed by a second Edward's Venn diagram (Supplementary Material C), where the highest number of haplotypes was shared among Irish *L. perenne* ecotypes and Southern European *L. perenne* ecotypes (14, Supplementary Material C). The values for Western European and Northern European groups were higher, each 10%, respectively.

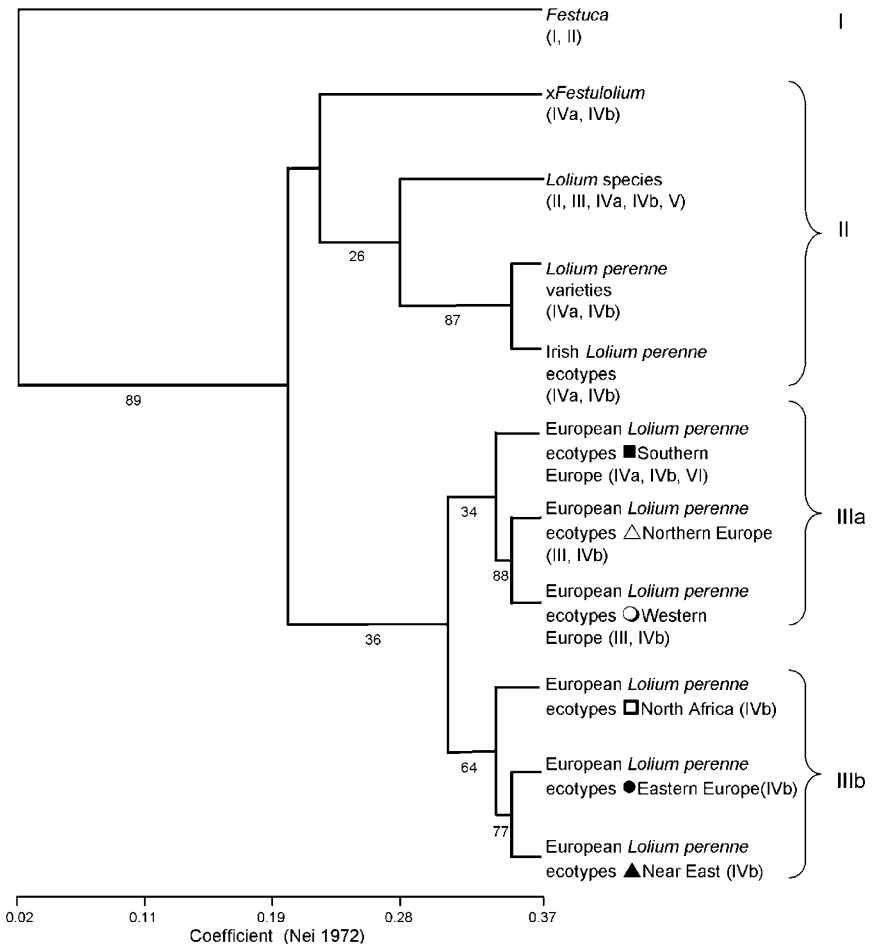

Figure 1 Unrooted dendrogram showing similarities between groups of accessions, constructed using the unweighted pair group method with arithmetic means (UPGMA) method (Sneath and Sokal, 1973) as implemented in NTSYSpc V2.2 (Rohlf, 2005), based on Nei's genetic distance measures (Nei, 1972). Numbers on the branches are percentage bootstrap values generated in NTSYSpc V2.2 (Rohlf, 2005). Symbols in parentheses are the groups where those species are present in Supplementary Material A.

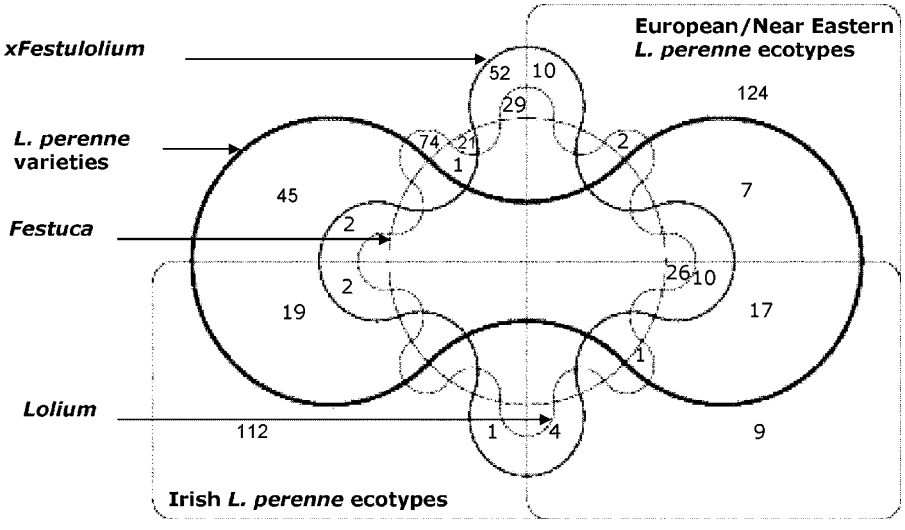

Figure 2 Edwards' Venn diagram demonstrating shared and unique haplotypes for the Irish *L. perenne* ecotypes, European/Near Eastern *L. perenne* ecotypes, *L. perenne* commercial varieties, other *Lolium* species, ×*Festulolium* varieties and *Festuca* species. Shared haplotypes among groups are in intersects.

Table 3 Diversity statistics based on Nei's (1987) analysis of gene diversity in subdivided populations for geographical groups of *L. perenne* ecotypes, *L. perenne* breeding varieties, *Lolium* species, × *Festulolium* varieties and *Festuca* species

	N_i	N_p	H_t	H_s	G_{st}
Irish *L. perenne* ecotypes	480	30	0.312	0.238	0.238
European and other geographic regional *L. perenne* ecotypes (total)	496	32	0.330	0.188	0.431
Northern Europe △	100	7	0.327	0.197	0.397
North Africa □	48	3	0.254	0.177	0.301
Near East ▲	60	4	0.233	0.146	0.373
Southern Europe ■	128	8	0.359	0.217	0.395
Western Europe ○	80	5	0.347	0.191	0.449
Eastern Europe •	80	5	0.247	0.166	0.328
L. perenne varieties	259	16	0.302	0.216	0.285
× *Festulolium*	140	9	0.318	0.227	0.285
Other *Lolium* species	136	11	0.341	0.142	0.585
Festuca species	64	6	0.269	0.077	0.716

Abbreviations: G_{st}, coefficient of genetic differentiation; H_s, diversity within subdivided populations; H_t, total gene diversity; N_i, number of individuals; N_p, number of populations.

Discussion

Characterization of cpDNA diversity at allelic and haplotypic level

All accessions displayed a high level of cpDNA SSR allelic variation and considerable numbers of haplotypes were found within ecotypes, within cultivars and within closely related groups of accessions. A total of 511 haplotypes were detected with an average of 10.4 haplotypes per accession. Partially this high variation within populations could be explained by the common and geographically widespread use of *L. perenne* as a cultivated agricultural species. Seed dispersal is the main factor affecting maternal plastid diversity over geographic space. Seeds could have been moved deliberately and accidentally by grazers, birds and wind or by human involvement including seed trade. We have detected numerous chloroplast haplotypes within *Lolium* populations, which would suggest that seed dispersal is high between populations over large geographical areas.

Some of the marker loci tested in our study were more variable than others. There are several possible reasons for this. For example, the locus with the least amount of variation, TeaCpSSR7, is located within a gene (*trn*H), whereas the locus with the highest amount of variation, TeaCpSSR3, is located in an intergenic spacer region (*trn*L-F intergenic spacer). This would be in accordance with the theoretical expectations for the evolution rate of these particular genomic regions (Wolfe *et al*., 1987). Moreover, the length of the cpSSR PCR product could play a role. At locus TeaCpSSR7, the length of the PCR product is shorter than the product at locus TeaCpSSR3 (McGrath *et al*., 2006). The length of the PCR product could reduce the theoretical possible likelihood of variation for a given length of sequence. Furthermore, longer cpSSR loci are known to display higher levels of molecular divergence than shorter cpSSR loci (Provan *et al*., 2001) partially because slipped-strand mispairing during DNA replication is greater within these regions.

Individual accessions showed a range of variation in gene diversity values, particularly among the ecotypes. For example, gene diversity values in Irish *L. perenne* ecotypes ranged from 0.122 to 0.333 (Table 1). Different factors acting on the individual populations may have affected the cytoplasmic diversity of the ecotypes. Isolation and fragmentation of individual populations could reduce diversity values, whereas increased movement of seed between certain ecotypes could have caused proportionally increased diversity.

The diversity of the Irish *L. perenne* ecotype populations was slightly less than in European *L. perenne* populations (H_t: 0.312 and 0.330, respectively). This could be explained in relation to the geographic position of Ireland as an island that has isolated Irish *L. perenne* ecotypes from the populations on the continent. If continental Europe was the centre of origin for *L. perenne*, the Irish diversity levels may be expected to be lower. However, plastid diversity is not markedly lower in Irish than in European *L. perenne* ecotypes and this could possibly be due to the thorough collection strategy of the Irish team in establishing the ECPGR collection that aimed to maximize the ecogeographical spread of Irish *L. perenne* populations. The G_{st} value of the Irish *L. perenne* ecotypes is almost half that of the European/Near Eastern *L. perenne* ecotypes (0.238 and 0.431, respectively). This indicates that the European/Near Eastern *L. perenne* ecotypes are more markedly different from each other than the Irish *L. perenne* ecotypes are from each other. This was not unexpected as there was often a larger geographic distance between European than between Irish populations. Because of this it may be argued that Irish ecotype accessions could be considered as a big meta-population. However, significant AMOVA variance components among Irish populations would contradict such a meta-population theory.

Haplotypes were also shown to be highly heterogenous within populations, with only 11 out of 104 populations containing no unique haplotypes (Table 1). This level of heterogeneity in populations indicated that seed for these populations was derived from many maternal lines, which is in accordance with breeding principles for allogamous forage species (Acquaah, 2006). Fifteen haplotypes were found in a study of 447 *L. perenne* and *L. rigidum* individuals (3%) by Balfourier *et al*. (2000), 41 haplotypes in a study of 168 *Fraxinus excelsior* individuals (24%) by Harbourne *et al*. (2005) compared to 511 haplotypes in 1575 individuals (32%) in this study. While 27% of haplotypes detected in the study by Balfourier *et al*. (2000) were unique to single populations, 71% of haplotypes detected in this study were unique to single populations. Our findings are in

Table 4 Analysis of molecular variance (AMOVA) for Irish and European *L. perenne* accessions, × *Festulolium* varieties and subgroups within European/Near Eastern *L. perenne* ecotype accessions

Source of variation	*Migration route*	*d.f.*	*SSD*	*Variance component*	*Variance (%)*	P[a]
Irish ecotypes	NA					
Among populations		29	140.17	0.23	18	***
Within populations		449	493.29	1.10	82	***
European ecotypes	NA					
Among populations		31	312.84	0.59	39	***
Within populations		462	427.57	0.93	61	***
Irish ecotypes vs European ecotypes	NA					
Among groups (Irish ecotypes vs European ecotypes)		1	95.24	0.18	11	***
Among populations/within groups		61	453.01	0.42	26	***
Within populations		911	920.86	1.01	63	***
Irish L. perenne ecotypes vs × Festulolium varieties	NA					
Among groups (Irish *L. perenne* ecotypes vs × *Festulolium* varieties)		1	74.98	0.32	19	***
Among populations/within groups		38	201.84	0.28	17	***
Within populations		580	622.31	1.07	64	***
European L. perenne ecotypes vs × Festulolium varieties	NA					
Among groups (European *L. perenne* ecotypes vs × *Festulolium* varieties)		1	47.00	0.17	10	***
Among populations/within groups		40	374.51	0.56	34	***
Within populations		593	556.59	0.94	56	***
Near East ▲ vs Southern Europe ■	Mediterranean, Danubian					
Among groups (Near Eastern ecotypes vs Southern European ecotypes)		1	8.95	0.00	−1	***
Southern Europe ■ vs Western Europe ○	Mediterranean					
Among groups (Southern European ecotypes vs Western European ecotypes)		1	11.76	0.00	0	***
Southern Europe ■ vs Eastern Europe •	Danubian					
Among groups (Southern European ecotypes vs Eastern European ecotypes)		1	18.48	0.06	4	***
Eastern Europe • vs Northern Europe Δ	Danubian					
Among groups (Eastern European ecotypes vs Northern European ecotypes)		1	12.92	0.04	3	***
Near East ▲ vs North Africa □	North African					
Among groups (Near Eastern ecotypes vs North African ecotypes)		1	10.01	0.03	2	***
North Africa □ vs Southern Europe ■	North African					
Among groups (North African ecotypes vs Southern European ecotypes)		1	15.65	0.08	4	***
All north of the alps ecotypes vs all south of the alps ecotypes	Post-glacial					
Among groups (All northern ecotypes vs all southern ecotypes)		1	10.63	0.00	0	***
Among populations/within groups		31	312.45	0.59	39	***
Within populations		477	444.82	0.93	61	NS
Southern Europe ■ vs Irish ecotypes	Into Ireland					
Among groups (Southern European ecotypes vs Irish ecotypes)		1	21.91	0.08	5	***
Among populations/within groups		36	216.10	0.31	21	***
Within populations		568	630.13	1.09	74	**
Western Europe ○ vs Irish ecotypes	Into Ireland					
Among groups (Western European ecotypes vs Irish ecotypes)		1	26.49	0.15	10	***
Among populations/within groups		33	192.76	0.30	19	***
Within populations		524	564.79	1.08	71	***
Northern Europe Δ vs Irish ecotypes	Into Ireland					
Among groups (Northern European ecotypes vs Irish ecotypes)		1	31.30	0.16	10	***
Among populations/within groups		35	198.72	0.29	19	***
Within populations		542	583.69	1.08	71	***

Abbreviations: d.f., degrees of freedom; NA, not applicable; NS, not significant; SSD, sum of squared differences.
[a]**$P<0.01$ and ***$P<0.001$.

accordance with findings of Fjellheim *et al*. (2006) in *Festuca* study including *L. perenne* for comparative reasons. Their analysis demonstrated that *L. perenne* accessions had four times as many haplotypes as *F. pratensis*. These authors suggested that a lower variation in *F. pratensis* could be due to the greater age of the species *F. pratensis* and because the chloroplast variation in *F. pratensis* was largely decimated in the Eemian interglacial period, where most of Europe was densely forested, while *F. pratensis* is a shade intolerant species.

Plastid genome identification, cytoplasmic gene pool characterization, and the study of hybridization and introgression

While none of the alleles for each of the ten cpSSRs were diagnostic for individual populations, several of the alleles found were unique to specific population groups (Table 2). It is possible that germplasm from these collections could be identified by genotyping these cpSSR markers if sufficient numbers of individuals are tested. Particularly useful for this purpose could be marker TeaCpSSR8, where half of the alleles were unique to Irish or European *L. perenne* ecotype accessions, respectively (Table 2). At the haplotype level, the majority of haplotypes detected were unique to specific groups of populations (Figure 2). While the high level of heterogeneity of haplotypes within populations made it difficult to assign individuals to specific populations, it was possible to use these haplotypes to assign individuals to specific groups. For example, 22% of haplotypes were specific to the Irish *L. perenne* ecotypes, and 24% of haplotypes were specific to the other European *L. perenne* ecotypes. These haplotypes have potential to distinguish geographic *L. perenne* ecotypes and accessions (Supplementary Material B).

The high level of variation, both allelic and haplotypic, in the European and Irish *L. perenne* ecotype collection in comparison with *L. perenne* cultivars suggests that the full cytoplasmic diversity is still underexploited in breeding material (Table 2 and Figure 2). Ecotypes with unique plastid variation not present in breeding material could be useful to expand the cytoplasmic gene pool for breeding of the species. A wide variation in plastid type can be useful to enhance the possibility of yield gains and yield stability as demonstrated on a data set for potato (Provan *et al*., 1999). For both UPGMA dendrograms (Supplementary Material A and Figure 1), the ten cpSSR markers were able to distinguish among Irish and European *L. perenne* ecotypes. This outcome indicated the usefulness of these ten cpSSR markers to identify cytoplasmic gene pools in ecotypes and breeding *L. perenne* germplasm.

Identification of plastid type is also useful for the study of introgression and hybridization (Hodkinson *et al*., 2002; Johannessen *et al*., 2005), as plastid marker information can identify the source of introgression and can be used in parentage analysis. For example, *L. temulentum* (L11) grouped with two *Festuca* species, NL1 and NL4 (Supplementary Material A, group II). This could be an indication of introgression of the plastid genome from *Festuca* species into *Lolium*. AMOVA analysis of Irish and European and Near Eastern *L. perenne* ecotypes vs the × *Festulolium* cultivars showed there was almost twice as much of the among-group variation between Irish *L. perenne* ecotypes and × *Festulolium* cultivars than between the European/Near Eastern *L. perenne* ecotypes and × *Festulolium* cultivars (Table 4). This could suggest more movement of cytoplasmic material between European ecotypes and × *Festulolium* cultivars than with Irish *L. perenne* ecotypes. Six out of nine × *Festulolium* cultivars grouped with the European *L. perenne* ecotypes (Supplementary Material A), which also could indicate introgression from European and Near Eastern *L. perenne* ecotypes into × *Festulolium*.

Plastid identification could also be used to verify that seed or seedlings derived from crosses in breeding programmes were assigned to the correct maternal parent (Gauthier *et al*., 1997). This could be particularly helpful for *Lolium* breeding in which multiple maternal lines are used in plant breeding (top cross breeding).

Phylogenetic and phylogeographic genetic structure of *Lolium*

Studying plastid DNA variation can contribute to phylogenetic analysis. UPGMA data demonstrated that two broad-leaved *Festuca* species, *F. arundinacea* and *F. pratensis*, and three narrow-leaved *Festuca species*, *F. pratensis*, *Festuca rubra* and *Festuca vivipara*, grouped together, respectively (Supplementary Material A). The broad-leaved *Festuca* species grouped closer to *Lolium*. Both of these findings were in agreement with previous studies (Darbyshire and Warwick, 1992; Catalan *et al*., 1997, 2004; Charmet *et al*., 1997; Torrecilla *et al*., 2004). However, some unusual groupings have occurred in the UPGMA dendrogram (Supplementary Material A). For example, one of the European *L. perenne* ecotypes (■17) was positioned outside all other *Lolium* accessions. This particular accession was from Sardinia where previously a high degree of diversity has been found for other species (Papa *et al*., 1998). Moreover, unlike other studies (Catalan *et al*., 1997, 2004; Charmet *et al*., 1997; Torrecilla *et al*., 2004), no separation of allogamous and autogamous *Lolium* species was found. These unusual groupings could be explained by high homoplasy in the data set caused by rapid molecular evolution at the loci studied. Parallel evolution at these loci would therefore be expected to be high and this would obscure phylogenetic signal of the markers (Flannery *et al*., 2006). They are thus potentially more useful for assessing variation within species than among species.

A loose correlation of genetic and geographic distances was detected with a Mantel test for the ecotypes were an exact geographic position was available ($r = 0.33$). While studies have tested the correlation between nuclear and geographic distances such as Cresswell *et al*. (2001), this is the first study to test the correlation between plastid genetic and geographic distances in *L. perenne*. We believe the lack of correlation is due to both the high within population plastid diversity and the high degree of seed-mediated gene flow, natural and human related.

The AMOVA analyses indicated that most of the variation in populations used in this study was within-groups and individual populations, but that there was also significant population genetic structuring among groups (Table 4). Generally, higher among-population variance component values are comparable to AMOVA

analysis results of other studies of *L. perenne* populations using nuclear markers (Bolaric *et al*., 2005).

The results for the AMOVA analysis showing the proportion of variance within or among groups were also useful for assessing broad-scale biogeographical patterns. For comparisons among groups examining the Mediterranean migration route in relation to other possible migration routes, the variance components between partitions for the Danubian and North African routes were 0 or close to 0. When a variance component is close to 0, it can mean that there is no population genetic structure (Schneider *et al*., 2000). Close to 0 or 0 values can also be an indication that samples among groups are more closely related to each other than samples within groups. This would indicate that these population groups are closer to each other than to groups showing a higher among-group variance component. For this data set it could be an indication of a Mediterranean movement of *L. perenne* from the Near East across Southern Europe into North Western Europe, Ireland. This is in accordance with one movement theory of *L. perenne* across Europe as proposed by Balfourier *et al*. (2000). This finding was substantiated in our study by the result of an AMOVA for post-glacial partitioning of south of the Alps ecotypes against north of the Alps ecotypes. In this case, among-group variation was 0 as well (Table 4). The post-glacial movement hypothesis can be further supported by our UPGMA data (Figure 1). Southern European, Northern European and Western European *L. perenne* ecotypes were grouped together and were distinct from the other European/Near Eastern *L. perenne* ecotypes. This indicated that these population groups were closer to each other than to other geographic groups and that movement of seed between these groups has occurred. Finally, the hypothesis that *L. perenne* most likely moved from the southern Europe into Ireland can be supported by the lowest among-group variation value in the AMOVA analysis (Table 2: 5%) for Southern Europe/Irish ecotype partition.

Conclusion

Allelic and haplotypic variation was extremely high within and between Irish and European *L. perenne* ecotypes. Migration of seed material by natural or anthropogenic means, including plant breeding, could contribute to this high level of variability. High plastid diversity was clearly persisting in populations. The cpSSR markers were shown to be extremely useful for characterizing variation in our accessions and have enabled the identification of cytoplasmic genepools and maternal lineages. The plastid type of individual populations could not be unambiguously identified, but groups of populations could be successfully identified. This suggests that an increase in the number of cpSSR markers would increase the likelihood of identifying individuals within population groups. Our findings describe broad-scale biogeographical patterns of population genetic structure in this highly heterogenous species. Furthermore, some evidence was provided to support possible broad-scale prehistorical geographical migrations. A pathway of migration from Southern Europe to Northwest Europe including Ireland is most likely.

Acknowledgements

SMcG is supported by a Teagasc PhD Walsh Fellowship. The Irish Department of Agriculture and Food supported the research with additional grant aid under the 'Genetic Resources for Food and Agriculture' scheme. We are grateful for discussions with Dr Vincent Connolly. Seeds were kindly provided by the genebank of the Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben/Germany, the National Genetic Resources Program (GRIN)/USA, Dr Radomir Capka of the Plant Breeding Station Hladke Zivotice/Czech Republic, Dr Alan Stewart of PGG Wrightson/New Zealand and Dr Ian Thomas of the genebank at International Grassland and Environmental Research (IGER)/Wales. We are also grateful for the excellent technical genotyping assistance of Ms Teresa Marie Charles.

References

Acquaah G (ed) (2006). Breeding cross-pollinated species. *Principles of Plant Genetics and Breeding*. Blackwell Publishing: Oxford. pp 313–333.

Balfourier F, Imbert C, Charmet G (2000). Evidence for the phylogeographic structure in *Lolium* species related to the spread of agriculture in Europe A cpDNA study. *Theor Appl Genet* **101**: 131–138.

Bolaric S, Barth S, Melchinger AE, Posselt UK (2005). Genetic diversity in European perennial ryegrass cultivars investigated with RAPD markers. *Plant Breed* **124**: 161–166.

Catalan P, Kellogg EA, Olmstead RG (1997). Phylogeny of Poaceae subfamily Pooideae based on chloroplast *ndh*F gene sequences. *Mol Phylogenet Evol* **8**: 150–166.

Catalan P, Torrecilla P, Lopez Rodriguez JA, Olmstead RG (2004). Phylogeny of the Festucoid grasses of subtribe Loliinae and allies (Pooeae, Pooideae) inferred from ITS and *trn*L-F sequences. *Mol Phylogenet Evol* **31**: 517–541.

Charmet G, Balfourier F, Chatard V (1996). Taxonomic relationships and interspecific hybridization in the genus *Lolium* (grasses). *Genet Resour Crop Evol* **43**: 319–327.

Charmet G, Ravel C, Balfourier F (1997). Phylogenetic analysis in the *Festuca–Lolium* complex using molecular markers and ITS rDNA. *Theor Appl Genet* **94**: 1038–1046.

Connolly V (2000). Collection and conservation of forage species from old pasture ecosystems. *Irish J Agric Food Res* **39**: 476–477.

Cresswell A, Sackville Hamilton NR, Roy AK, Viegas MF (2001). Use of amplified fragment length polymorphism markers to assess genetic diversity of *Lolium* species from Portugal. *Mol Ecol* **10**: 229–241.

Daniell H, Kumar S, Dufourmantel N (2005). Breakthrough in chloroplast genetic engineering of agronomically important crops. *Trends Biotechnol* **23**: 238–245.

Darbyshire SJ, Warwick SI (1992). Phylogeny of North American *Festuca* (Poaceae) and related genera using chloroplast DNA restriction site variation. *Can J Bot* **70**: 2415–2429.

Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.

Fjellheim S, Rognli OA, Fosnes K, Brochmann C (2006). Phylogeographical history of the widespread meadow fescue (*Festuca pratensis* Huds.) inferred from chloroplast DNA sequences. *J Biogeograph* **33**: 1470–1478.

Flannery ML, Mitchell FJG, Coyne S, Kavanagh TA, Burke JI, Salamin N *et al* (2006). Plastid genome characterisation in *Brassica* and Brassicaceae using a new set of nine SSRs. *Theor Appl Genet* **113**: 1221–1231.

Gauthier P, Lumaret R, Bedcarrats A (1997). Chloroplast-DNA variation in the genus *Lotus* (Fabaceae) and further evidence regarding the maternal parentage of *Lotus corniculatus* L. *Theor Appl Genet* **95**: 629–636.

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**: 463–471.

Gray JC (2005). Regulation of nuclear gene expression by plastid signals. In: Moller SG (ed). *Plastids*. Blackwell Publishing: Oxford. pp 237–267.

Hallden C, Gertsson B, Sall T, Lind-Hallden C (1993). Characterisation of organellar DNA in alloplasmic lines of *Brassica napus* L. *Plant Breed* **111**: 185–191.

Harbourne ME, Douglas GC, Waldren S, Hodkinson TR (2005). Characterization and primer development for amplification of chloroplast microsatellite regions of *Fraxinus excelsior. J Plant Res* **118***: 339–341.*

Hodkinson TR, Chase MW, Lledo MD, Salamin N, Renvoize SA (2002). Phylogenetics of *Miscanthus*, *Saccharum* and related genera (Saccharinae, Andropogonae, Poaceae) based on DNA sequences from ITS nuclear ribosomal DNA and plastid *trn*L intron and *trn*L-F intergenic spacers. *J Plant Res* **115**: 381–392.

Huang SX, Sirikhachornkit A, Faris JD, Su X, Gill BS, Haselkorn R *et al* (2002). Phylogenetic analysis of the acetyl-CoA carboxylase and 3-phosphoglycerate kinase loci in wheat and other grasses. *Plant Mol Biol* **48**: 805–820.

Humphreys MW, Yadav RS, Cairns AJ, Turner LB, Humphreys J, Skøt L (2006). A changing climate for grassland research. *New Phytologist* **169**: 9–26.

Johannessen MM, Andersen BA, Damgaard C, Jorgensen RB (2005). Maternal inheritance of chloroplasts between *Brassica rapa* and F1-hybrids demonstrated by cpDNA markers specific to oilseed rape and *B. rapa*. *Mol Breed* **16**: 271–278.

Joshi SP, Ranjekar PK, Gupta VS (1999). Molecular markers in plant genome analysis. *Curr Sci* **77**: 230–240.

McGrath S, Hodkinson TR, Salamin N, Barth S (2006). Development and testing of novel chloroplast microsatellite markers for *Lolium perenne* and other grasses (Poaceae) from *de novo* sequencing and *in silico* sequences. *Mol Ecol Notes* **6**: 449–452.

Michalakis Y, Excoffier L (1996). A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* **142**: 1061–1064.

Nei M (1972). Genetic distance between populations. *Am Nat* **106**: 283–292.

Nei M (1973). Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* **70**: 3321–3323.

Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press: New York.

Papa R, Attene G, Barcaccia G, Ohgata A, Konishi T (1998). Genetic diversity in landrace populations of Hordeum vulgare L. from Sardinia, Italy, as revealed by RAPDs, isozymes and morphophenological traits. *Plant Breeding* **117**: 523–530.

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995). Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proc Natl Acad Sci USA* **92**: 7759–7763.

Provan J, Powell W, Dewar H, Bryan G, Machray GC, Waugh R (1999). An extreme cytoplasmic bottleneck in the modern European cultivated potato (*Solanum tuberosum*) is not reflected in decreased levels of nuclear diversity. *Proc R Soc Lond B Biol Sci* **266**: 633–639.

Provan J, Powell W, Hollingsworth PM (2001). Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol Evol* **16**: 142–147.

Rohlf F (2005). *NTSYSpc: Numerical Taxonomy System*, Version 2.2 (2). Exeter Publishing Ltd: Setauket, NY.

Saski C, Lee SB, Daniell H, Wood TC, Tomkins J, Kim HG *et al.* (2005). Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. *Plant Mol Biol* **59**: 309–322.

Schneider S, Roessli D, Excoffier L (2000). *Arlequin: A Software for Population Genetics Data Analysis*. Version 2.000. Genetics and Biometry Lab, Department of Anthropology, University of Geneva.

Sneath P, Sokal R (1973). *Numerical Taxonomy.* W.H. Freeman and Company: San Francisco.

Torrecilla P, Lopez Rodriguez JA, Catalan P (2004). Phylogenetic relationships of *Vulpia* and related genera (Poeae, Poaceae) based on analysis of ITS and *trn*L-F sequences. *Ann Miss Bot Gard* **91**: 124–158.

United Nations Publication (1999). *Standard Country or Area Codes for Statistical Use* (Department of Economic and Social Affairs, United Nations) Revision 4. url: http://millenniumindicators.un.org/unsd/methods/m49/m49.htm.

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

Wolfe KH, Li W-H, Sharp PM (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA* **84**: 9054–9058.

Yeh FC, Boyle T (1997). Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Bel J Bot* **129**: 157.

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