

Genetic clues to the origin of the apple

Stephen A. Harris, Julian P. Robinson and Barrie E. Juniper

Molecular genetic markers complement archaeological, breeding and geographical investigations of the origins, history and domestication of plants. With increasing access to wild apples from Central Asia, along with the use of molecular genetic markers capable of distinguishing between species, and explicit methods of phylogeny reconstruction, it is now possible to test hypotheses about the origin of the domesticated apple. Analyses of nuclear rDNA and chloroplast DNA (cpDNA) sequences indicate that the domesticated apple is most closely related to series *Malus* species. Moreover, the occurrence of a shared 18-bp duplication in the cpDNAs of wild and cultivated apple supports the close relationship between them. Hypotheses about the hybridization and the origin of the domesticated apple cannot be rejected completely until more variable, phylogenetically informative markers are found.

The domesticated apple is one of the most important fruit crops of the colder and temperate parts of the world [1]. Vavilov suggested that the wild apple of Turkestan (Kazakhstan, Kyrgystan, Uzbekistan, Turkmenistan and Tajikistan) and its close relatives were the progenitors of the domesticated apple*; the whole process of wild apple domestication being traced to Almaty (Kazakhstan). Vavilov reasoned that, because the wild apple bears similar fruits to the domesticated apple, it must have been the progenitor. Recent fieldwork in the region appears to confirm the similarity between wild and cultivated apples [2,3]. Furthermore, Janick *et al.* suggested that 'this area [Central Asia] is the area of greatest diversity and the centre of origin' of the domesticated apple [4]. This is linked to Vavilov's 'oversimplified' idea that the centre of diversity is the place of origin [5,6]. The Central Asian wild apple (Box 1) is a diverse species with a wide range of forms, colours and flavours [7], whilst its allozyme diversity is significantly greater than that

* Vavilov, N.I. (1930) Wild progenitors of the fruit trees of Turkistan and the Caucasus and the problem of the origin of fruit trees. *International Horticultural Congress Group B*, 271–286

Box 1. Taxonomic hierarchy and apple classification

Species are arranged according to a taxonomic hierarchy, in which they are placed into larger groups or divided into smaller groups to account for observed variation. Each of these groups is given a particular name. Thus, the family Rosaceae includes the genus *Malus* (apples) with its ~55 species, which are grouped into infrageneric groups (section, series), and each species can be divided into intraspecific groups (cultivar). *Malus* spp. are arranged according to Phipps *et al.*'s classification [a].

The genus *Malus* comprises some 55 species [a], although between eight and 79 species have been recognized. Part of the problem of *Malus* taxonomy is the intimate association that humans have with apples, where the distinction between wild and cultivated species could become blurred and hence the recognition of distinct categories difficult. Furthermore, the scientific names that have been applied to the domesticated apple are legion, and include *Malus pumila* Miller, *M. communis* Desf., *M. sylvestris* (L.) Miller and *M. domestica* Borkh. Recently, it has been argued that *M. pumila* is the correct name for the domesticated apple and that this also includes the presumed wild relative, *M. sieversii* (Ledeb.) M. Roem [b]. We use *M. domestica* to refer to 'domesticated apple'

and *M. sieversii* to refer to the Central Asian 'wild apple'. However, the term 'wild apple' has been applied to all *Malus* spp. other than the domesticated apple. For example, in Europe, the common name 'wild apple' refers to *M. sylvestris* Miller, whilst in North America the term has been applied to species including *M. angustifolia* (Ait.) Michx. and *M. coronaria* (L.) Miller. The situation is further complicated by the application of the name 'wild apple' to hybrids between domesticated apple and other members of the genus *Malus*.

The taxonomic hierarchy for the domesticated apple 'Bramley's Seedling' would be:

- **Family:** Rosaceae
- **Subfamily:** Maloideae
- **Genus:** *Malus*
- **Section:** *Malus*
- **Series:** *Malus*
- **Species:** *domestica*
- **Variety/cultivar:** 'Bramley's seedling'

References

- a Phipps, J.B. *et al.* (1990) A checklist of the subfamily Maloideae (Rosaceae). *Can. J. Bot.* 68, 2209–2269
- b Mabberley, D.J. *et al.* (2001) The name of the apple. *Telopea* 9, 421–430

found in four widely distributed North American wild apples [8,9].

The Central Asian wild apple is closely related to a group of apples that have different-sized fruits. One of these, *Malus baccata* (Siberian Crab), has small, red fruits that hang in clusters and are bird-dispersed (Fig. 1). *Malus baccata* might have had a wider distribution than that of the present day, and we think that populations became 'trapped' as the Tien Shan began to rise out of the Tethys Ocean. Over seven million years, perhaps

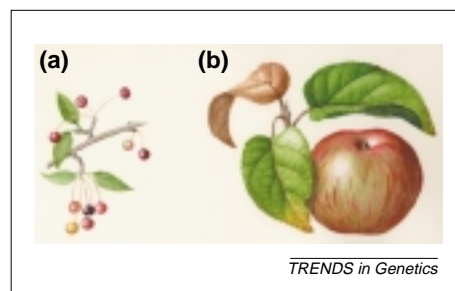


Fig. 1. Range of variation in apple fruit types: (a) *Malus baccata*, a bird-dispersed species; and (b) domesticated apple, a mammal-dispersed species.

up to ten million years, mammals, such as bears, acted as distribution vehicles by selecting the largest and juiciest fruits; a small, bird-distributed, cherry-like delicacy giving way to a large mammal-distributed form. Small apples have even been observed to pass intact through a bear's jaw and guts; it should be noted, however, that apple seeds retained in the apple core do not germinate.

By the time humans began to occupy the area ~5000–8000 years ago, the early evolution of the apple was almost complete, and its migration, ably assisted by the now domesticated horse [10], was under way. Over several more thousands of years, from within this migrating flow came the many thousands of apple cultivars now known, as a result of both unconscious and conscious selection [11].

Based on combined archaeological and molecular data, it seems likely that, in the late Neolithic or early Bronze Age, travellers on the great trade routes that ran from central China to the Danube (Fig. 2), carried the seed of the Central

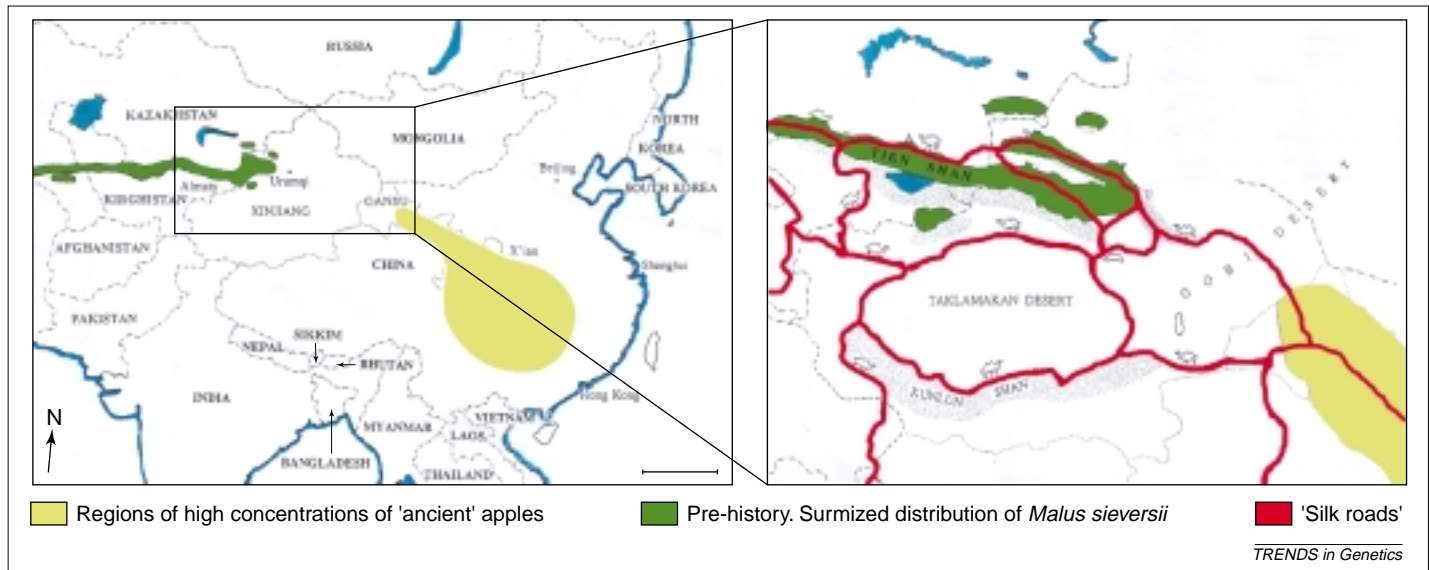


Fig. 2. Location of great trade routes ('silk roads', red), surmized distribution of *Malus sieversii* (green) and regions of high concentrations of present-day 'ancient' apples (yellow). Scale bar = 500 km.

Asian wild apple west, either in saddle bags or horses' guts. We now know that the important technique of grafting (described as 'instant domestication' [12]), where the variety can be preserved forever, was probably discovered in Mesopotamia at Mari, as early as 3800 years ago (S. Dalley, pers. commun.). The description, on cuneiform tablets, concerns vines (*Vitis* spp.), but the techniques are easily transferable. From there, the fruits and the necessary technology passed through the Persians and Greeks to the Romans, who perfected orchard economies. The Romans brought the whole package to Western Europe and, for the last 2000 years, the domesticated apple has diversified and flourished worldwide. Archaeological evidence for the collection of apples from the wild in Europe can be found in Neolithic (11 200 years ago) and Bronze Age (c. 4500 years ago) sites throughout Europe [13,14], and for apple cultivation as early as 1000 BC in Israel [1].

In terms of apple domestication, series *Malus* is the most important group of species. However, the nomenclature of the species in series *Malus* is complex. With few discrete characters to differentiate species, the difficulty of species delimitation has hampered investigations into the origins of apples [15]; morphological characters used to delimit species in series *Malus* are continuous and overlapping. Such problems perhaps contributed to Zohary and Hopf's statement that 'it is therefore futile to try to delimit the area of initial domestication on

the basis of the evidence available from the living plant' [1]. However, the molecular systematics revolution potentially provides an excellent source of potential characters for analyses of these type.

The origin of the domesticated apple and relations with other *Malus* spp.

Combinations of nuclear (nDNA) and cytoplasmically [chloroplast (cpDNA) and mitochondrion (mtDNA)] inherited molecular genetic markers provide important clues about the origins of domesticated plants, for example, *Citrus* spp. [16], beans [17] and potatoes [18]. In *Malus*, the chloroplast genome provides data about evolutionary relationships of the maternal line, whilst biparentally inherited nDNA provides independent data, which, combined with cpDNA, could enable the origin of the domesticated apple to be determined. Hybridization and introgression are expected to have been important factors in the development of domesticated apple cultivars. Therefore, for investigations that aim to determine the earliest origins of the domesticated apple, it is necessary to use cultivars that are as ancient as possible.

matK, a cpDNA-encoded region ~1800-bp long, of which 1341 bp was sequenced, showed only 16 phylogenetically informative characters across the genus *Malus*, and thus poor resolution in the phylogenetic tree [19] (Fig. 3). However, two duplications were found 39 bp from the 3' end of the *matK* coding region. Duplication I is an imperfect

8-bp duplication that differs by a T residue, whereas duplication II is a perfect 18-bp duplication. Duplications I and II were always present together, suggesting that each has arisen only once during the evolution of *Malus* spp. (Fig. 3).

Duplication II was polymorphic in both the Central Asian wild apple and the domesticated apple, suggesting that the Central Asian wild apple could be the major maternal contributor, as proposed by Watkins *et al.* [20].

The nuclear ribosomal internal transcribed spacer (ITS) showed 89 phylogenetically informative characters from the 617 bp sequenced. A strongly supported group comprised the Central Asian wild apple and the domesticated apple, as well as *M. asiatica*, *M. orientalis*, *M. niedzwetzkyana* and *M. prunifolia* (all from section *Malus* spp.), although there was little resolution within this group (Fig. 4). Together, data from the chloroplast and nuclear genomes support the view that the domesticated apple is most closely related to the series *Malus* spp. These data also indicate that the Central Asian wild apple could be most closely related to the domesticated apple. Savolainen *et al.* were unable to determine the origin of the domesticated apple because of a lack of variation in the cpDNA-encoded *atpB-rbcL* spacer and the absence of the Central Asian wild apple in their sample [21]. However, additional phylogenetically useful DNA sequence information should be sought, so that this hypothesis can be confirmed; additional sampling of Central Asian wild apples and domesticated apples is needed to ensure that rare hybridization events with other *Malus* spp. that might have

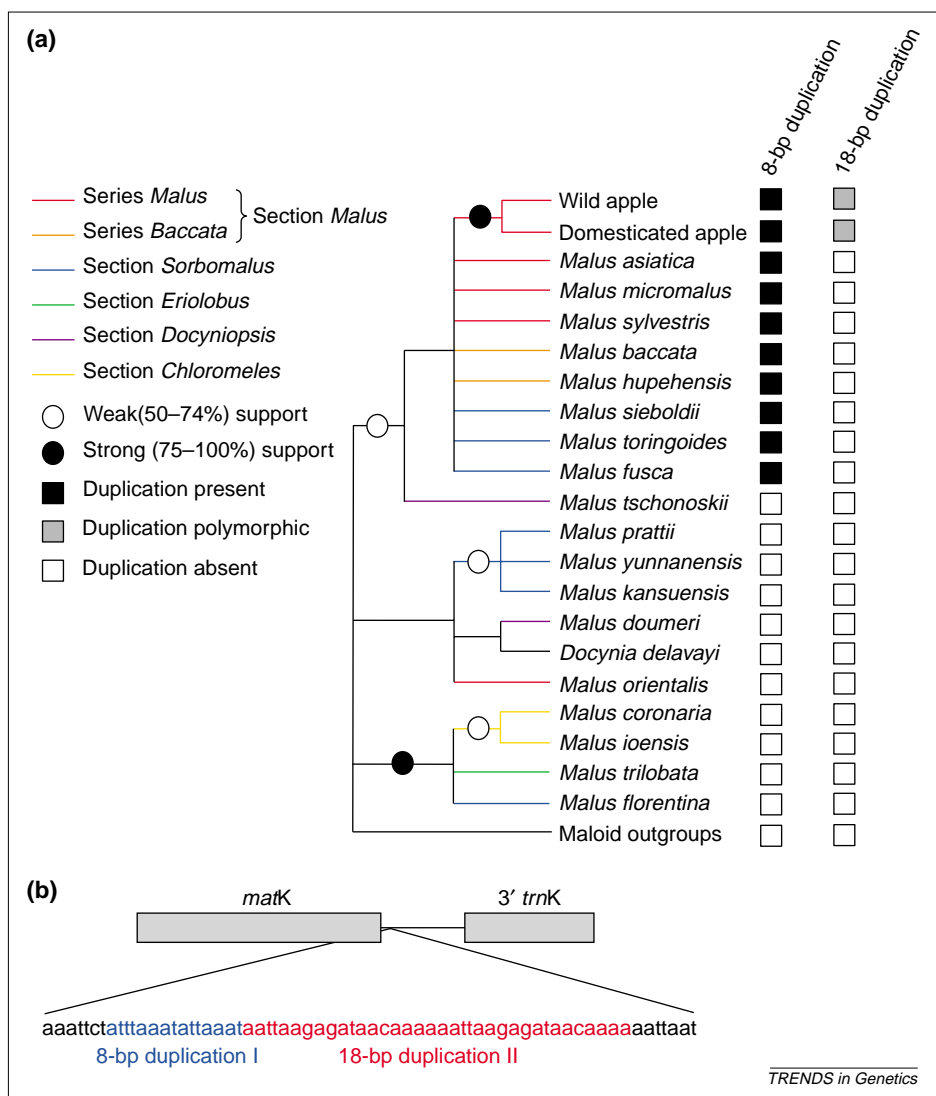


Fig. 3. Phylogeny of the chloroplast DNA-encoded gene *matK* in the genus *Malus*. (a) A cartoon of *matK* is shown in (b) and the positions of the two duplications are indicated. Wild apple refers to the Central Asian wild apple, *M. sieversii*. Circles indicate the level of statistical support for particular groups. The distribution of the two *matK* duplications is shown on the tree.

contributed to the early domestication of apples have not been overlooked.

The need for more variable DNA markers has led to analyses of *Malus* phylogeny using randomly amplified polymorphic DNA (RAPD) [22] and nuclear microsatellites [single sequence repeats (SSR)] [23]. Dunemann *et al.*'s results indicated that *M. pumila* and *M. sylvestris* (section *Malus*) were involved in the origin of cultivated apples [23]. However, they suggested that *M. sylvestris*, *M. florentina* or *M. dasycphylla* could be the female parent of *M. domestica*. By contrast, Hokanson *et al.*'s SSR studies were not useful in determining the relationships among the 142 accessions from 23 species analysed [23]. However, it has been argued that RAPD data are not appropriate for the reconstruction of phylogeny because of

reproducibility, primer structure, dominance, product competition, homology, allelic variation, genome sampling and non-independence of loci, whilst similar problems have been highlighted for the use of SSRs in phylogenetic analyses [23–25].

Variation within the domesticated apple

Many thousands of named domestic apple varieties (both dessert and cider) have been selected for hundreds of years in Europe, Asia and North America, and more recently in the Southern Hemisphere. Together with wild species, these are maintained in national collections as genetic resources for breeding, particularly as sources of resistance to apple scab [*Venturia inaequalis* (Cooke) Winter], powdery mildew [*Podosphaera leucotricha* (Ellis and Everh) Salmon] and fireblight

[*Erwinia amylovora* (Burill) Winslow *et al.*] [26]. As with wild species, the taxonomy of apple varieties is fraught with problems, including environmental and developmental variation within varieties, and limited numbers of morphological and chemical characters to distinguish varieties [11]. In addition the triad of decreasing budgets, but increasing collection size and running cost, traps a collection manager whose goal is to have a well-characterized but easily utilized collection, and forces choices to be made to try and minimize genotypic redundancy. Thus, molecular genetic markers provide a valuable aid to the identification of duplicates (possible cultivar synonyms), correctly naming mis-identified plants and enabling managers to make effective decisions about the reduction of collection size without reducing genetic variation (i.e. creation of core collections) [27].

Two contrasting types of molecular genetic marker can be used to characterize apples. The first type is those that identify few, high information content loci, such as microsatellites. The second are those that identify many, low information content loci, such as RAPD. Studies of RAPD variation in *Malus* spp. have focused on the diversity present in domesticated apples and germplasm collections. Dunemann *et al.* investigated RAPD variation at 52 putative loci in a collection of 27 domesticated apple cultivars and were able to differentiate between cultivars [22]. Furthermore, RAPD markers have also been used to analyse the maternal and paternal contributions to pedigrees [22,28,29]. In a large study by Oraguzie *et al.*, 43 RAPD markers were used to assess the variation in a worldwide collection of 50 old domesticated apple varieties, current breeding lines and a collection of 105 other apples, including wild species [30]. Individual cultivars could be distinguished and ~95% of the RAPD variation occurred within cultivars, breeding lines and wild species. Furthermore, Central Asian wild apple accessions were scattered in a cluster analysis based on genetic distance. However, RAPD markers have been criticized for the analysis of genetic variation on both technical and theoretical grounds [24] and, given the problem of poor inter-laboratory result reproduction, arbitrary fragment length polymorphisms could be a more effective way of producing genotypes based on many, low information content, loci [31].

the University of Tashkent for assistance in the field collecting in Uzbekistan and Kirghistan; the staff of LES-IC, the Kirghiz Swiss Forestry Support Programme and the Forestry Institute of the Republic of Kirghistan, Jalal-Abad, Kirghistan; N. Ajtkhozina and N. Ajtkhozina of the Ministry of Sciences and the University of Almaty for assistance in field collecting in Kazakhstan. In particular, we thank R. Wise of the Dept of Plant Sciences, Oxford, UK for all the artwork. B.E.J. is currently a Leverhulme Emeritus Research Fellow in the Dept of Plant Sciences, Oxford.

References

- Zohary, D. and Hopf, M. (2000) *Domestication of Plants in the Old World*, Oxford University Press
- Forsline, P.L. (1995) Adding diversity to the national apple germplasm collection: collecting wild apples in Kazakhstan. *N. Y. Fruit Quart.* 3, 3–6
- Forsline, P.L. *et al.* (1994) Collection of wild *Malus*, *Vitis* and other fruit species genetic resources in Kazakhstan and neighboring republics. *HortScience* 29, 433
- Janick, J. *et al.* (1996) Apples. In *Fruit Breeding: Tree and Tropical Fruits* (Janick, J. and Moore, J.N., eds), pp. 1–77, John Wiley & Sons
- Vavilov, N.I. (1926) Studies on the origin of cultivated plants. *Trudy Byuro. Prikl. Bot.* 16, 139–245
- Harlan, J.R. (1992) *Crops and Man*, American Society of Agronomy
- Way, R.D. *et al.* (1990) Apples (*Malus*). In *Genetic Resources of Temperate Fruit and Nuts* (Moore, J.N. and Ballington, R., eds), pp. 1–62, International Society of Horticultural Science
- Lamboy, W.F. *et al.* (1996) Partitioning of allozyme diversity in wild populations of *Malus sieversii* L. and implications for germplasm collection. *J. Am. Soc. Hort. Sci.* 121, 982–987
- Dickson, E.E. *et al.* (1991) Isozymes in North American *Malus* (Rosaceae): hybridisation and species differentiation. *Syst. Bot.* 16, 363–375
- Vila, C. *et al.* (2001) Widespread origin of domestic horse lineages. *Science* 291, 474–477
- Morgan, J. (1993) *The Book of Apples*, Ebony Press
- Zohary, D. and Spiegel-Roy, P. (1975) Beginnings of fruit growing in the Old World. *Science* 187, 319–327
- Hopf, M. (1973) Apfel (*Malus communis* L.); Aprikose (*Prunus armeniaca* L.). In *Reallexikon der Germanischen Altertumskunde* (Vol. 1) (Beck, H. *et al.* eds), pp. 363–372, Walter de Gruyter
- Schweiggruber, F.H. (1979) Wildapfel und prähistorische Apfel. *Archaeo-Physika* 8, 283–294
- Rohrer, J.R. *et al.* (1994) Floral morphology of Maloideae (Rosaceae) and its systematic relevance. *Am. J. Bot.* 81, 574–581
- Moore, G.A. (2001) Oranges and lemons: clues to the taxonomy of *Citrus* from molecular markers. *Trends Genet.* 17, 536–540
- Beccara Velásquez, V.L. and Gepts, P. (1994) RFLP diversity of common bean (*Phaseolus vulgaris*) in its centres of origin. *Genome* 37, 256–263
- Hosaka, K. (1995) Successive domestication and evolution of the Andean potatoes as revealed by chloroplast DNA restriction endonuclease analysis. *Theor. Appl. Genet.* 90, 356–363
- Robinson, J. *et al.* (2001) Taxonomy of the genus *Malus* Mill. (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh. *Plant Syst. Evol.* 226, 35–38
- Watkins, R. (1995) Apple and pear. In *Evolution of Crop Plants* (Smatt, J. and Simmonds, N.W., eds), pp. 418–422, Longman
- Savolainen, V. *et al.* (1995) Chloroplast DNA variation and parentage analysis in 55 apples. *Theor. Appl. Genet.* 90, 1138–1141
- Dunemann, F. *et al.* (1994) Genetic relationships in *Malus* evaluated by RAPD 'fingerprinting' of cultivars and wild species. *Plant Breed.* 113, 150–159
- Hokanson, S.C. *et al.* (2001) Microsatellite (SSR) variation in a collection of *Malus* (apple) species and hybrids. *Euphytica* 118, 281–294
- Harris, S.A. (1999) RAPDs in systematics – a useful methodology? In *Advances in Molecular Systematics* (Hollingsworth, P.M., *et al.* eds), pp. 211–228, Taylor & Francis
- Robinson, J. and Harris, S.A. (2000) Amplified fragment length polymorphisms and microsatellites: a phylogenetic perspective. In *Which DNA Marker for Which Purpose?* (Gillet, E.M., ed.), pp. 95–121, Institut für Forstgenetik und Forstpflanzenzüchtung
- Manganaris, A.G. *et al.* (1994) Isozyme locus *Pgm-1* is tightly linked to a gene (V_r) for scab resistance in apple. *J. Am. Soc. Hort. Sci.* 119, 1286–1288
- Szewc-McFadden, A.K. *et al.* (1996) Utilisation of identified simple sequence repeats (SSRs) in *Malus x domestica* (apple) for germplasm characterisation. *HortScience* 31, 619
- Harada, T. *et al.* (1993) DNA-RAPDs detect genetic variation and paternity in *Malus Euphytica* 65, 87–91
- Landry, B.S. *et al.* (1994) Phylogeny analysis of 25 apple rootstocks using RAPD markers and tactical gene tagging. *Theor. Appl. Genet.* 89, 847–852
- Oraguzie Nnadozie, C. *et al.* (2001) Genetic diversity and relationships in *Malus* sp. germplasm collections as determined by randomly amplified polymorphic DNA. *J. Am. Soc. Hort. Sci.* 126, 318–328
- Jones, C.J. *et al.* (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Mol. Breed.* 3, 381–390
- Szewc-McFadden, A.K. *et al.* (1995) Identification of simple sequence repeats in *Malus* (apple). *HortScience* 30, 855
- Hokanson, S.C. *et al.* (1998) Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus x domestica* Borkh. core subset collection. *Theor. Appl. Genet.* 97, 671–683
- Ishikawa, S. *et al.* (1992) Organelle DNA polymorphism in apple cultivars and rootstocks. *Theor. Appl. Genet.* 83, 963–967
- Kato, S. *et al.* (1993) Mitochondrial DNA restriction fragment length polymorphisms in *Malus* species. *Plant Breed.* 111, 162–165

Stephen A. Harris

Barrie E. Juniper*

Dept of Plant Sciences, University of Oxford, South Parks Road, Oxford, UK OX1 3RB.

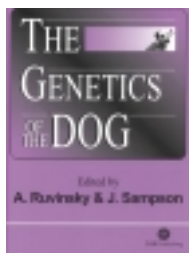
*e-mail: barrie.juniper@plants.ox.ac.uk

Julian P. Robinson

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, UK OX1 3RE.

Book Review

Dog-gone good genetics



The Genetics of the Dog
edited by A. Ruvinsky and J. Sampson
CAB International, 2001. £85.00 hbk
(X + 564 pages)
ISBN 0851995209

Phenotypic variation is the stuff of geneticists' dreams. With hundreds of pure breeds showing extraordinary

differences in size, shape, physiology and behavior, the domestic dog is uniquely suited among mammals for modern genetic analysis. Why then has canine genetics not had a more prominent role in this, the genomics era? There are several reasons, some of which become apparent while reading *The Genetics of the Dog*. This book is a loosely organized collection of reports from leading canine geneticists that describe the latest research results and molecular tools. Unfortunately, this is all the book aspires to do. *The Genetics of the Dog* does not address obstacles in the field or put forth hypotheses that catalyze

future research. As is often the case with multi-authored compendiums, there is little continuity. The chapters are not coordinated to support a common context or theme, leading to a lack of overall commentary on the 'big picture' of canine genetics. The book also fails to create a sense of excitement and discovery around a species that has so much to offer to studies of mammalian development, behavior, evolution and health. Thus, this book represents a missed opportunity – to be provocative, to chart a course for future work, and to excite readers with the enormous promise that dog genetics