Chapter 13
Establishing Chromosome Genomics in Forage and Turf Grasses

D. Kopecký, J. Číhalíková, J. Kopecká, J. Vrána, M. Havránková, Š. Stočes, J. Bartoš, H. Šimková, J. Šafář, M. Kubaláková, P. Navrátil and J. Doležel

Abstract Analyses of large genomes are hampered by high proportions of repetitive DNA, that make the assembly of short sequence reads difficult. This is also the case for meadow fescue (*Festuca pratensis* Huds.), one of predominant grass species in temperate and Northern regions with the genome size estimated at 1C = 3,175 Mbp. This species is known for its ability to survive under freezing conditions and it has been used widely in intergeneric hybridization with various ryegrass species to produce superior Festulolium cultivars. Here we describe attempts to dissect the meadow fescue's genome into smaller fractions—individual chromosomes and groups of chromosomes. Following the methods of flow cytogenetics developed for legumes and cereals, we have developed a chromosome sorting protocol for grasses and currently we are able to sort *F. pratensis* chromosome 4 (the largest in the genome) and two groups of three chromosomes each: 2, 3, 7 and 1, 5, 6. As the first step we sequenced chromosome 4 by Illumina with 50x coverage and assembled low copy and genic regions. This facilitated detailed comparative analysis with sequenced genomes of rice, *Brachypodium* and sorghum and provided the first insight into the genome composition of this species. The possibility to purify chromosome 4 opens the way for a more efficient analysis of genetic loci on this chromosome that control important agronomic traits, such as freezing tolerance. Moreover, purified chromosomes are excellent templates for PCR screening as well as cytogenetic and physical mapping.

13.1 Introduction

Ryegrasses and fescues are among the predominant forage and turf species in temperate climates. Ryegrass species are known for their high yield, palatability, digestibility, rapid establishment, favorable nutrient characteristics, dark green color and uniformity of turf. However, they are susceptible to stress conditions such as

D. Kopecký (✉) · J. Číhalíková · J. Kopecká · J. Vrána · M. Havránková · Š. Stočes · J. Bartoš · H. Šimková · J. Šafář · M. Kubaláková · P. Navrátil · J. Doležel
Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Sokolovská 6, 77200 Olomouc, Czech Republic
e-mail: kopecky@ueb.cas.cz

S. Barth, D. Milbourne (eds.), *Breeding Strategies for Sustainable Forage and Turf Grass Improvement,*
DOI 10.1007/978-94-007-4555-1_13,

winter freezing, and heat and drought during summer. Fescues have complementary characteristics to those of ryegrasses. Therefore, they are widely combined in mixtures. They have also become subjects of intergeneric hybridization. Tens of x Festulolium cultivars with various genomic constitutions have been released and are widely used in forage and turf seed industry in Europe (Kopecký et al. 2006).

Despite their agronomic importance, the progress in genetics, genomics and bioinformatics of grasses is far behind that of cereals. Among others, such studies are complicated by the outcrossing nature of these species, population-based breeding and frequent aneuploidy (Kopecky et al. 2005). Moreover, genomes of species within *Festuca-Lolium* are large and complex. The genome of meadow fescue was estimated at 1C = 3175 Mbp based on flow-cytometry, similar to that of human (Doležel et al. 2003; Kopecký et al. 2010). Ryegrass genome is just a little smaller (1C = 2623 Mbp of *L. perenne* and 1C = 2567 Mbp of *L. multiflorum*). Because of the complexity and size of genomes (full of repeats), any procedural or technical simplification would be highly welcome in genomic studies.

13.2 Chromosome Genomics

Analysis of a complex genome where a majority of DNA is present as repeats can be simplified by several approaches. Several methods were used to avoid sequencing of the repetitive parts of genomes. Sequencing of cDNAs to generate ESTs (Expressed Sequence Tags) is among the most successful. However, ESTs fail to sample rare or conditional transcripts (Martienssen et al. 2004) and other methods were proposed to target the gene space in large and complex genomes (Cot fractionation and methyl filtration) (Peterson et al. 2002; Rabinowicz et al. 1999). Unfortunately, these two methods did not provide the expected improvements.

Another alternative to overcome the complexity of large genomes is to dissect them to smaller parts and sequencing these parts individually. Working with naturally uniform and independent units—chromosomes—is perhaps the most powerful approach. In hexaploid wheat, individual chromosomes represent only 1–3 % of the entire genome (Doležel et al. 2009) and even diploid species' chromosomes (as meadow fescue) dissect the 3.2 Gbp genome into 373–543 Mbp units, each representing 11.7–17.1 % of the whole (Kopecký et al. 2010).

There are two ways to isolate individual chromosomes: microdissection and flow-sorting. Microdissection enables the isolation of any chromosome or a chromosome segment. However, the total yield is limited to only a few copies of a particular chromosome (Zhou and Hu 2007). Flow cytometric sorting relies on differences in chromosome size. It can generate samples of unlimited copy numbers of specific chromosomes with purity usually exceeding 90 %. Moreover, DNA of sorted chromosomes is suitable for further molecular analysis (Doležel et al. 2004). The output of flow cytometry analysis is a histogram of relative chromosome fluorescence intensity (reflecting chromosome size), which is called a flow karyotype (Fig. 13.1). Ideally, each chromosome is represented by a single peak on a flow karyotype. However,

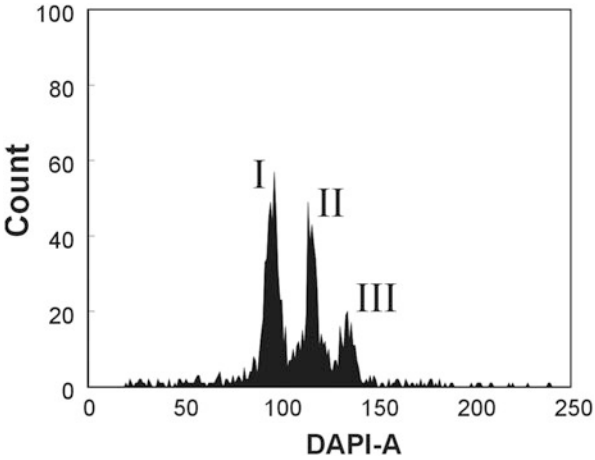

Fig. 13.1 Flow karyotyping in *F. pratensis* ($2n = 2x = 14$). The flow karyotype consists of two peaks representing groups of chromosomes and one peak representing chromosome 4F. Peak 1 represents small chromosomes 1F, 5F and 6F and Peak 2 consists of chromosomes 2F, 3F and 7F. *X axis* relative DAPI fluorescence intensity, *Y axis* number of events

chromosome size similarities frequently result in the formation of composite peaks formed by mixtures of two or more chromosomes. Doležel et al. (2009) proposed that there has to be at least 10 % difference in chromosome size to generate a separate peak on a flow karyotype. Unfortunately, in most plant species chromosomes are not that different. For example, flow cytometry analysis of hexaploid wheat (21 chromosome pairs) produced only four peaks, and only one of those contained a single chromosome, 3B (Vrána et al. 2000). All other chromosomes were present in composite peaks and could not be sorted into uniform fractions. Similarly, in barley only chromosome 1H, and in rye only chromosome 1R can be sorted into pure fractions (Suchánková et al. 2006; Kubaláková et al. 2003). However, the plasticity of plant genomes (especially in polyploids) makes it possible to develop special cytogenetic stocks with reconstructed karyotypes. Telosomic lines made it possible to sort individual chromosome arms in wheat; in instances where a telocentric overlaps in size with smaller complete chromosomes, isochromosomes provide a solution. Similarly, wheat-rye and wheat-barley addition lines enabled sorting of individual chromosomes or chromosome arms of barley and rye (Doležel et al. 2009). Nowadays, all wheat, barley and rye chromosomes/chromosome arms can be isolated in pure fractions and are being used for physical mapping, sequencing and map-based cloning.

13.3 Karyotypes of Grasses (Fescues and Ryegrasses)

Meadow fescue (*Festuca pratensis* Huds.) is a diploid species with seven chromosome pairs. The largest is metacentric chromosome 4, followed by another metacentric chromosome 3 (the one with a prominent secondary constriction), submetacentric chromosome 2 and metacentric chromosome 7. The smallest are submetacentric chromosomes 5, 6 and 1. Individual lengths (as determined at C-metaphase) vary

from 4.67 μm to 6.79 μm (Kopecký et al. 2010). Italian ryegrass (*Lolium multiflorum* Lam.) has a similar karyotype, with chromosome length ranging from 3.37 μm (chromosome 1) to 5.33 μm of chromosome 4 (Kopecký et al. 2010). The largest is the metacentric chromosome 4. Similar in length are submetacentric chromosome 2 and metacentric chromosomes 3 and 7. These three chromosomes carry secondary constrictions. The smallest are submetacentric chromosomes 5, 6 and 1. Perennial ryegrass (*Lolium perenne* L.) has a very similar karyotype. On the other hand, the karyotype of hexaploid tall fescue (*F. arundinacea* Schreb.) consists of 21 chromosome pairs of similar size.

The variation in length of diploid grass species' chromosomes was promising for separation (and sorting) of individual chromosomes using flow cytometry. The above mentioned over 10 % difference in chromosome size allowed us to predict that chromosome 4 would form separate peak on the flow karyotype and generate a pure sample, while the remaining six chromosomes would form two composite peaks.

13.4 Flow Cytometry and Chromosome Genomics in Grasses

After overcoming problems with the synchronization of cell cycle in such puny little root tips of germinating seed of meadow fescue, we were able to isolate and stain mitotic chromosomes for a flow cytometry analysis. As predicted, the flow karyotype of meadow fescue consists of three peaks (Fig. 13.1). The first (composite) peak represents three smallest chromosomes—1F, 5F and 6F. Another three chromosomes (2F, 3F and 7F) form the composite second peak. The third peak represents chromosome 4F. Thus, this chromosome is the only one which can be directly sorted from the standard karyotype. Nowadays, we routinely sort chromosome 4F with purity exceeding 92 %.

The two composite peaks of the meadow fescue flow karyotype may be partitioned into sections, as done in wheat, theoretically enabling sorting of particular chromosomes with a reasonable purity. Moreover, special cytogenetic stocks could be used to sort specific chromosome constructs, such as those used in cereals. The difference in chromosome size between *L. multiflorum* and *F. pratensis* warrants a prediction that chromosomes 2F, 3F and 7F can be sorted from disomic substitution lines, where two homologous chromosomes of tetraploid *L. multiflorum* are substituted by their homoeologues from *F. pratensis* (Kopecký et al. 2008a). We do have such lines available for every chromosome of *F. pratensis*.

13.5 Applications for Sorted Grass Chromosomes

Sorted chromosomes can be used for a wide range of applications in genomics, genetics and cytogenetics.

13.5.1 Bacterial Artificial Chromosome (BAC) Libraries

The development of a chromosome specific BAC library is one of the most attractive uses of flow-sorted chromosomes. The strategy of fingerprinting a well defined part of the genome (particularly single chromosomes/chromosome arms) and sequencing clone by clone is certainly more precise and efficient than the whole genome sequencing, especially in all cases where high proportions of repeats make contig assembly a challenging proposition. Chromosome specific BAC libraries are attractive resources for physical mapping by ordering BAC clones according to their fingerprint patterns. Such libraries have been developed for most of wheat chromosomes/chromosome arms and the short arm of rye chromosome 1 (1RS) (Šimková et al. 2008; Šafář et al. 2010). The utility of chromosome specific BAC libraries was demonstrated by the construction of a physical map of chromosome 3B of wheat (Paux et al. 2008).

Until now, only the whole genome BAC libraries have been available for species of the *Festuca—Lolium* complex. Donnison et al. (2005) and Farrar et al. (2007) developed BAC libraries for *F. pratensis* and *L. perenne*, respectively. A partial BAC library for *F. pratensis* was developed and used for cytogenetic mapping by Kopecký et al. (2008b, 2010).

Successful sorting of chromosome 4 of *Festuca* (and presumably of *Lolium*) opens the way for the development of a BAC library specific for this chromosome, BAC libraries for chromosomes 2F, 3F and 7F could be potentially created using single chromosome substitution lines. Given the above statement on partitioning of composite peaks, we can assume that enriched BAC libraries for other chromosomes (with purity of ~65 %) could be developed as well.

13.5.2 Development of Molecular Markers

Chromosome specific BAC libraries are a valuable source of molecular markers. By BAC End Sequencing (BES), Paux et al. (2008) generated over 700 ISBP (Insertion Site-Based Polymorphisms) markers for chromosome 3B, which are already used in breeding programs. Similarly, using BES, Bartoš et al. (2008) and Kofler et al. (2008) developed ISBP and SSR markers, respectively, specific for the short arm of chromosome 1R of rye. The advantage of chromosome-based approaches for marker development is that the chromosomal specificity of markers can be verified by PCR on flow-sorted chromosomes serving as templates.

Genetic maps are available for both ryegrass species (*L. multiflorum* and *L. perenne*), as well as for tall and meadow fescues. However, the genetic map of meadow fescue comprises only 466 markers (Alm et al. 2003). The development of the DArTFest array (Kopecký et al. 2009) enriched the genetic map by additional almost 150 DArT markers (Bartoš et al. 2011), however, the number of markers is still limiting for fine genetic mapping and positional cloning of agronomically important genes. Thus, a low number of molecular markers for meadow fescue calls for the

use of chromosome-based approach. Besides the development of markers by BAC-end sequencing, the example of wheat chromosome 3B shows that a combination of flow sorting and the DArT technology is fully capable of yielding high numbers of chromosome specific DArT markers (Wenzl et al. 2010).

13.5.3 Cytogenetic Mapping Using Sorted Chromosomes

Sorted chromosomes can be used as an alternative template for cytogenetic mapping, instead of traditional cytological preps squashed or dropped on a microscopic slide. The advantage of this approach is the purity of a chromosome fraction on a slide. Moreover, dropped individual chromosomes solve the interference problem brought about by the presence of cytoplasm and cell wall residues always present on squashed preparations. Spatial resolution and sensitivity of the technique can also be significantly improved by using super stretched sorted chromosomes (Valárik et al. 2004).

We used sorted chromosomes as a template for cytogenetic mapping of various DNA sequences in *F. pratensis*. Localization of several clones from a partial BAC library identified all seven chromosomes of *F. pratensis*. The detailed cytogenetic mapping of various sequences (BAC clones and repeats) on chromosome 4 showed the effectiveness of this method. In total, we localized five BAC clones and 13 repeats (centromeric and telomeric repeats and 11 microsatellites). Surprisingly, the telomeric repeat originating from wheat produced signals not only in the telomeric regions of both chromosome arms, but also in a proximal region of the long arm (4FL). Similarly, additional signals were also detected in intercalary regions of chromosome arms 2FL, 5FL and 6FS (where S and L denote short and long arms, respectively). These signals might represent positions of ancient structural rearrangements. Signals in the centromeric/pericentromeric region was also detected by FISH with BAC clones 1F21, 1G18 (additional signals on both arms), 2B14 (additional signal on the short arm), 2D4 (additional signal on the long arm) and 2N9 (additional signals on both arms). Trials with 11 microsatellites (motifs CA, GC, TA, CAG, CAT, CGG, GAA, GAC, GAG, TAA and TAC) on chromosome 4F did not provide satisfactory results. Only the microsatellite with the CAG motif provided a signal in the distal part of the short arm of 4F. No other microsatellites were detected on this chromosome.

13.5.4 Sequencing of Individual Chromosomes

Sequencing of BAC ends from a chromosome specific BAC library provides an insight into the molecular organization of individual chromosomes. By sequencing ends of almost 11,000 BAC clones of wheat chromosome 3B-specific library, Paux et al. (2008) has shown that 86 % of sequences were repeats and only 1.2 % came from coding regions. The balance were sequences of unknown role or function. Similar

results were obtained by sequencing ends of rye 1RS specific BAC clones (Bartoš et al. 2008).

Nowadays, the dramatic reduction in cost and increase in effectiveness of the Next Generation Sequencing enables sequencing of entire chromosomes (and even genomes) in a reasonable time and for a reasonable cost (Doležel et al. 2009). Mayer et al. (2011) sequenced individual chromosomes of barley using the Roche 454 technology with 1.28–2.76x coverage. This enables the prediction of total number of barley genes (~32,000) and provided a tool for fine comparative studies with model species: rice, sorghum and Brachypodium. Moreover, a virtual gene order has been designed for all seven barley chromosomes.

We sequenced chromosome 4F by Illumina HiSeq2000 with the coverage of over 50x. Using the same approach of comparative analysis as Mayer et al. (2011) and Wicker et al. (2011), we were able to identify collinear regions of *F. pratensis* chromosome 4 in genomes of barley and three model species: rice, sorghum and Brachypodium.

Acknowledgments Authors would like to thank Prof. Adam J. Lukaszewski for critical reading and valuable comments.This work has been supported by the Czech Science Foundation (grant award P501/11/0504) and by the Ministry of Education, Youth and Sports of the Czech Republic and the European Regional Development Fund (Operational Programme Research and Development for Innovations No. CZ.1.05/2.1.00/01.0007).

References

Alm V, Fang C, Busso CS, Devos KM, Vollan K, Grieg Z, Rognli OA (2003) A linkage map of meadow fescue (*Festuca pratensis* Huds.). Theor Appl Genet 108:25–40

Bartoš J, Paux E, Kofler R, Havránková M, Kopecký D, Suchánková P, Šafář J, Šimková H, Town CD, Lelley T, Feuillet C, Doležel J (2008) A first survey of the rye (*Secale cereale*) genome composition through BAC end sequencing of the short arm of chromosome 1R. BMC Plant Biol 8:95

Bartoš J, Sandve SR, Kölliker R, Kopecký D, Christelová P, Stočes S, Østrem L, Larsen A, Kilian A, Rognli OA and Doležel J (2011) Genetic mapping of DArT markers in the *Festuca-Lolium* complex and their use in freezing tolerance association analysis. Theor Appl Genet 122:1133–1147

Doležel J, Bartoš J, Voglmayr H, Greilhuber J (2003) Nuclear DNA content and genome size of trout and human. Cytometry 51:127–128

Doležel J, Kubaláková M, Vrána J, Bartoš J (2004) Flow cytogenetics. In: Goodman RM (ed) Encyclopedia of plant and crop science, Marcel Dekker, Inc., New York, pp 460–463

Doležel J, Šimková H, Kubaláková M, Šafář J, Suchánková P, Číhalíková J, Bartoš J, Valárik M (2009) Chromosome Genomics in the Triticeae. In: Feuillet C, Muehlbauer GJ (eds) Genetics and genomics of the triticeae, plant genetics and genomics: crops and models 7. Springer, Heidelberg, pp 285–316

Donnison IS, O'Sullivan DM, Thomas A, Canter P, Moore B, Armstead I, Thomas H, Edwards KJ, King IP (2005) Construction of a *Festuca pratensis* BAC library for map-based cloning in festulolium substitution lines. Theor Appl Genet 110:846–851

Farrar K, Asp T, Lubberstedt T, Xu M, Thomas AM, Christiansen C, Humphreys MO, Donnison IS (2007) Construction of two *Lolium perenne* BAC libraries and identification of BACs containing candidate genes for disease resistance and forage quality. Mol Breeding 19:15–23

Kofler R, Bartoš J, Gong L, Stift G, Suchánková P, Šimková H, Berenyi M, Burg K, Doležel J, Lelley T (2008) Development of microsatellite markers specific for the short arm of rye (*Secale cereale* L.) chromosome 1. Theor Appl Genet 117:915–926

Kopecky D, Lukaszewski AJ, Gibeault V (2005) Reduction of ploidy level by androgenesis in intergeneric *Lolium-Festuca* hybrids for turf grass breeding. Crop Sci 45:274–281

Kopecký D, Loureiro J, Zwierzykowski Z, Ghesquiere M, Dolezel J (2006) Genome constitution and evolution in *Lolium* × *Festuca* hybrid cultivars (Festulolium). Theor Appl Genet 113:731–742

Kopecký D, Lukaszewski AJ, Doležel J (2008a) Meiotic behaviour of individual chromosomes of *Festuca pratensis* in tetraploid *Lolium multiflorum*. Chromosome Res 16:987–998

Kopecký D, Lukaszewski AJ, Doležel J (2008b) Cytogenetics of Festulolium (*Festuca* × *Lolium* hybrids). Cytogenet Genome Res 120:370–383

Kopecký D, Bartoš J, Lukaszewski AJ, Baird JH, Černoch V, Kölliker R, Rognli OA, Blois H, Caig V, Lübberstedt T, Studer B, Shaw P, Doležel J, Kilian A (2009) Development and mapping of DArT markers within the *Festuca–Lolium* complex. BMC Genomics 10:473

Kopecký D, Havránková M, Loureiro J, Castro S, Lukaszewski AJ, Bartoš J, Kopecká J, Doležel J (2010) Physical distribution of homoeologous recombination in individual chromosomes of *Lolium multiflorum/Festuca pratensis* substitutions. Cytogenet Genome Res 129:162–172

Kubaláková M, Valárik M, Bartoš J, Vrána J, Číhalíková J, Molnár-Lang M, Doležel J (2003) Analysis and sorting of rye (*Secale cereale* L.) chromosomes using flow cytometry. Genome 46:893–905

Martienssen RA, Rabinowicz PD, O'Shaughnessy A, McCombie WR (2004) Sequencing the maize genome. Curr Opin in Plant Biol 7:102–107

Mayer KFX, Martis M, Hedley PE, Šimková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubaláková M, Suchánková P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Doležel J, Waugh R, Stein N (2011) Unlocking the Barley Genome by Chromosomal and Comparative Genomics. Plant Cell 23:1249–1263

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeyer W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S, Berges H, Eversole K, Appels R, Šafář J, Šimková H, Doležel J, Bernard M, Feuillet C (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. Sci 322:101–104

Peterson DG, Wessler SR, Paterson AH (2002) Efficient capture of unique sequences from eukaryotic genomes. Trends Genet 18:547–550

Rabinowicz PD, Schutz K, Dedhia N, Yordan C, Pernell LD, Stein L, McCombie WR, Martienssen RA (1999) Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. Nat Genet 23:305–308

Šafář J, Šimková H, Kubaláková M, Číhalíková J, Suchánková P, Bartoš J, Doležel J (2010) Development of chromosome-specific BAC resources for genomics of bread wheat. Cytogenet Genome Res 129:211–223

Šimková H, Šafář J, Suchánková P, Kovářová P, Bartoš J, Kubaláková M, Janda J, Číhalíková J, Mago R, Lelley T, Doležel J (2008) A novel resource for genomics of Triticeae: BAC library specific for the short arm of rye (*Secale cereale* L.) chromosome 1R (1RS). BMC Genomics 9:237

Suchánková P, Kubaláková M, Kovářová P, Bartoš J, Číhalíková J, Molnár-Lang M, Endo TR, Doležel J (2006) Dissection of the nuclear genome of barley by chromosome flow sorting. Theor Appl Genet 113:651–659

Valárik M, Bartoš J, Kovářová P, Kubaláková M, de Jong JH, Doležel J (2004) High-resolution FISH on super-stretched flow-sorted plant chromosomes. Plant J 37:940–950

Vrána J, Kubaláková M, Šimková H, Číhalíková J, Lysák MA, Doležel J (2000) Flow-sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). Genetics 156:2033–2041

Wenzl P, Suchánková P, Carling J, Šimková H, Huttner E, Kubaláková M, Sourdille P, Paul E, Feuillet C, Kilian A, Doležel J (2010) Isolated chromosomes as a new and efficient source of DArT markers for the saturation of genetic maps. Theor Appl Genet 121:465–474

Wicker T, Mayer KFX, Gundlach H, Martis M, Steuernagel B, Scholz U, Šimková H, Kubaláková M, Choulet F, Taudien S, Platzer M, Feuillet C, Fahima T, Budak H, Doležel J, Keller B, Stein N (2011) Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. Plant Cell 23:1706–1718.

Zhou RN and Hu ZM (2007) The development of chromosome microdissection and microcloning technique and its applications in genomic research. Curr Genomics 8:67–72