Searching and mapping genes involved in adaptation and speciation: revision, constraints and future orientations.

Mayra Sáenz Oviedo

Los Andes University. Bogotá D. C., Colombia.

The understanding of the identity of genes involved in speciation and adaptation is of vital importance in evolutionary biology inasmuch as the knowledge of the genetic basis of reproductive isolation and of the phenotypic variability, is fundamental in order to understand the processes of evolutionary relevance that generate diversification and adaptation. So far, and through mapping techniques, it has been possible an important approach in the search of such loci in organism like *Tribolium castaneum* (red flour beetle) *Bombyx mori* (silk worm), *Colias eurytheme*, *C. philodice*, *Heliconius melpomene* and *H. erato* (passion vine butterflies) among others. However, there are few studies that so far have achieved the isolation of genes involved in adaptation and speciation in organisms of biological interest. Among those there are *Drosophila* (fruit fly), *Coereba flaveola* (bananaquits), *Xiphophorus* (platyfish) and *Gasteosteus aculeatus* (three-spined sticklebacks). Because of this, and in response to the need of a much finer dissection in the search of the genes involved in such processes, new experimental alternatives with greater genome coverage and sensibility in the establishment of genetic associations with the traits of interest should be explored. Some of those methodologies that could accomplish the former characteristics are considered here, such as 454 sequencing RADs (Restriction associated DNA) and SNPs (Single nucleotide polymorphism), which could improve the identification of relevant loci in speciation and adaptation.

Adaptation and speciation genes

Until recently, the genes responsible of speciation and adaptation had remained unknown, leaving most of the evolutionary questions about the identity and characteristics of these genes unresolved. However, with the advance of some classic techniques and the development of several new techniques, some authors have achieved the isolation of genes involved in speciation and adaptation (see Box 3). Among these genes, those of hybrid sterility and inviability in *Drosophila* (Presgraves, 2003; Orr, *et al*., 2004) and fishes of the *Xiphophorus* genre (Orr *et al*., 2004); genes that cause polymorphism in the plumage of C*oereba flaveola* (bananaquits; Theron *et al*., 2001), and genes that control the number of bone plates in different morphs of *Gasterosteus aculeatus* (three-spined sticklebacks; Colosimo *et al*., 2005), studies that I will refer later.

The identification of the genes involved in adaptation and speciation is of vital importance in evolutionary biology, just like the knowledge of the molecular basis of the phenotypic diversity within the species, since the adaptive morphological evolution depends on the selection of genetic variants that may benefit processes like the colonization of new niches and the development of mechanisms of reproductive isolation (Orr *et al*., 2004; Theron *et al*., 2001; Colosimo *et al*., 2005). In addition, a better understanding of the molecular basis of processes involved in the parallel evolution requires the identification of genes and mutations that lead the phenotypic change (Orr *et al*., 2004; Presgraves, 2003).

With this review I will make a revision of the classic mapping techniques, their application in the order Insecta, and in the isolation of genes involved in speciation and adaptation. Moreover, to present the multiple limitations of such techniques and to propose new methodologies that may offer a greater genome coverage and sensitivity in the establishment of genetic associations with traits of interest.

Mapping and generation of molecular markers

In order to determine the evolutionary history of organisms of interest, besides of the molecular basis of the processes of adaptation and speciation, genome mapping has been widely used, which searches the creation of a genetic map assigning fragments of DNA to chromosomes through the use of markers that may be products of PCR (Polymerase chain reaction) or of restriction enzymes. The posterior ordering of these fragments, to generate the maps, it is derived from the recombination frequencies of the different genetic markers or from the use of computational programs that integrate the data of the banding profile generated by the restriction enzymes (Hui, 1998).

Within the classic techniques used to map traits, there are the AFLP (Amplified Fragment Length Polymorphism), SCNL (Single Copy Nuclear Loci), RFLP

(Restriction Fragment Length Polymorphism), RAPD (Restriction Amplified Polymorphic DNA) and STR (Short Tandem Repeats) (see Box1).

Through these techniques it has been possible to generate linkage maps for different groups of Insecta (see Box2). Additionally, with the advance of these and the development of several new techniques, it has been possible to isolate adaptation and speciation genes (see Box 3).

The linkage maps have been used to the identification of chromosome regions that contain genes that control simple traits (controlled by one single gene) and quantitative traits using Quantitative Trait Loci (QTL) analysis (Hui, 1998; Collard *et al*., 2005). In this way, the mapping of genes constitutes a key tool to decode the molecular basis of the phenotypic variation.

There are several categories of mapping, the genetic or linkage mapping, which is based on the recombination during meiosis (Hui, 1998); and the physical mapping, which derives from the assembly of contigs (a cluster of fragments of a genome that have been cloned individually, that are adjacent and partially overlapped; Hartl & Jones, 2005). Next, I will look into them.

In addition, the mapping process (genetic and physic) is based on the number of molecular markers. As a consequence, the greater the number of markers, the greater the genome coverage. The molecular markers are extensively used to define specific genes through genetic localization, for making associations between genomic location and function and for exploring the genetic variation during evolution (Hui, 1998; Collard *et al*., 2005).

Genetic mapping

The genetic map of an animal or plant species is an abstract model of a lineal arrangement of a group of genes and markers, providing the position and distances between markers along the chromosomes (Hui, 1998; Collard *et al*.,

2005). In this way, it can be seen like a road map where the molecular markers are the signals.

This type of map is base don the homologous recombination during meiosis, so it can be called a meiotic map as well (Hui, 1998). This suggests that if two or more markers are located near on a chromosome, their alleles are usually inherited together through meiosis (Hui, 1998; Futuyma, 2005; Hartl & Jones, 2005). This characteristic allows the analysis of genes and markers in the progeny (Collard *et al*., 2005).

For the construction of a genetic map it is necessary having a set of data composed of a large number of segregant genetic markers (Hui, 1998; Collard *et al*., 2005). Furthermore, the construction of a linkage map requires of three main steps (Fig. 1a): (i) the production of a mapping population, derived from sexual reproduction and composed of more than 100 individuals; (ii) it is necessary to identify the polymorphisms, recognizing the molecular markers that reveal the differences between the parents and that subsequently are screened on the mapping population; finally (iii) a linkage analysis is performed through programs like MapMaker (Lander *et al*., 1987), that involves the codification of data for each marker in each individual of the population, analyzing and determining the linkage between large numbers of markers, which is calculated with a logarithm of odds (LOD) which is defined as the logarithm of the rate of linkage versus no linkage (Collard *et al*., 2005)

The typical result of a linkage map (Fig. 1b), should have markers distributed uniformly throughout the genome, in order to posses a complete representation of it (with no bias towards one or another region), and in this way be able to perform future analysis like those of QTL. Likewise, it should contain anchor loci that could be found in other maps, so it is possible to make comparative studies along the genome with other organisms of interest and evaluate synteny and the conserved order of the genes through orders, families or species (Collard *et al*., 2005; Yasukochi *et al*., 2006).

Physical mapping

A physical map is another mechanism to make associations between traits and DNA sequences. Commonly, these maps contain arranged fragments of cloned DNA in different vectors and that are obtained through the digestion with restriction enzymes. A restriction map in a common form of a physical map of a chromosomal segment, but the most extreme example of a physical map is the DNA sequence (Hui, 1998).

The physical mapping usually are elaborated from gene libraries in which the genome is fragmented in multiple random and disorganized pieces, each one of them cloned separately in an adequate vector. The idea of elaborating a physical map consists of the ordering of these genome fragments with the purpose of finding sets of pairs of fragments partially overlapped or contigs. In this way, the current physical maps assemble the fragments in a contig, and afterwards, the different contigs that belong to the same linkage group are assembled, describing in detail the position and physical distances (in base pairs) of and between different markers (Avise, 2004; Hui, 1998; Iañez, E., 1997; Fig. 2).

This category of maps is available with greater frequency for some animals and plants species with big chromosomes and extensive citogenetic information. However, due to the nature of its technique that assigns genes to determined locations along a chromosome using measurements that are a reflection of the physical distance between the genes (Thompson *et al*., 2004), the physical maps usually have low resolution in comparison with the genetic maps. In this way a gene can be frequently located to a specific chromosome and commonly to a region within such chromosome, but the resolution on a DNA level is low (Hui, 1998).

Constraints and future orientations

Despite the significative advance of the former mentioned techniques, there are restrictions in their application; among which, the low reproducibility (of the

RAPDs) and the high costs and difficult laboratory methods (like the elaboration of the STRs primers) are the most important.

Constraints

Among the main constraints of the classic techniques, there is the restriction in the number of markers and of their combinations, in that the resolution of the genetic analysis depends in great measure of these, because the greater the number of markers, the greater the genome coverage, and a better estimation of the genetic distances. Another limitation is the mapping of traits that are difficult to quantify, like behavior or mating preference in which the segregation is not easily distinguishable, making the study of this kind of phenomena quite demanding. Furthermore, the QTL statistics are rather complicated to develop, in such way that the employment of these is not possible for all the organisms of study. Additionally there is a problem of resolution, because most of the times these techniques do not allow the identification and isolation of particular genes, but frequently are limited to finding regions or intervals where the gene of interest is most likely located. Besides, there is the problem of the scale, because for example, for humans one centimorgan (cM) equals one million of base pairs (Mbp), which makes it difficult to know for other organisms of interest what is the correspondence of the map units, and therefore to know what is the real resolution the is being obtained. Finally, these classic techniques of polymorphic molecular markers generation are somewhat laborious and expensive which decreases their applicability.

Future orientations

Recently, different solutions have been introduced in order to overcome the preceding restriction and therefore advance in the genomic study. Firstly, there is the implementation of mapping techniques with RADs, SNPs and 454 sequencing, which I will refer to later.

In addition, to this implementation, it is advisable to obtain a confirmation that the gene found by mapping techniques, is actually generating the expected effect, which could be done by means of gene interference or blocking.

454 Pyrosequencing

This new technique of fast sequencing that involves a parallel massive amplification through a PCR emulsion and the analysis of sequences over optic fiber chips (Fig. 2), presents multiple benefits, between which Ellegren (2008) cited:

1. It gives direct access to information of sequences useful for approaches for candidate genes in regions that show signals of linkage to phenotypic traits.
2. It is a valuable source of genetic markers for QTL or association mapping, and population genetics analysis due to the great amount of sequences variants that can be detected.
3. With the sequences information availability at great scale of the transcriptome (the part of the genome that is transcript to RNA), species-specific microarrays can be constructed in order to make further comparative studies.

It is important to establish that the acquisition of substantial numbers of contigs does not imply that and organism has a large quantity of genes, because many of the sequences found by the sequencing of cDNA or EST do not represent genes that codify for proteins, for example, they could be RNAs that do not codify. In this way, the scientifics use all the unique sequences obtained in BLAST quests against other genomes with the intention of having a first approach to the genome of interest (Ellegren, 2008).

Thanks to this technique, Vera *et al*. (2008) carried out a study with the butterfly *Melitaea cinxia*, where two runs of 454 were performed, and 48000 contigs formed by two or more overlapping readings and 60000 singletons were assembled: most of the singletons and probably also the contigs were originated form the same gene, even when it is not possible in every opportunity

to assemble all of them in a continuous sequence. Finally, a BLAST search against the genome of *Bombyx mori* was carried out, and 9000 coincidences were found. Taking into account that *B.mori* has around 18000 genes and that this is representative of Lepidoptera, it could be affirmed that at least half of the genome of this butterfly has already been mapped at least in part.

RAD (Restriction Associated DNA)

The DNA tags associated to restriction sites (RAD) are a representation through short DNA labels and throughout the genome of every restriction site of a particular restriction enzyme. Most of the organisms segregate high numbers of DNA polymorphism that interrupt the restriction sites, which allow the RAD tags to be employed as genetic markers that are in high densities along the genome. Additionally, it has applicability for genotyping of individuals and for bulk segregant analysis (Miller *et al*., 2007).

In this way, this technique provides an approach that permits the screening of almost all the enzyme restriction sites in parallel. Like this, this method can identify and type a great number of markers and it is a resource easy to produce in model and non-model organisms. RAD markers can be identified by detecting differential patterns of hybridization on samples labeled with RAD on a microarray (Miller *et al*., 2008; Fig. 3).

SNP (Single Nucleotide Polymorphism)

The polymorphism of one single nucleotide (SNP) exhibit the type of molecular variation more common in the genome, in such way that they constitute really good molecular markers because of their high density and uniform distribution throughout the genome in codifying and non-codifying regions (which permits linkage within genes sequences), representing ideal candidates for studying the heritability in genomic regions (Li *et al*., 2007; Hoskins *et al*., 2008).

Hoskins *et al* (2008), with these markers mapped *GLUT9*; a gene located in the fourth human chromosome that codifies for a glucose transporter and that is

mainly expressed in the kidney and liver. They were able to find that the SNP rs6855911 was in strong association with this gene, in order to reveal that the polymorphisms found in this gene can alter the glucose metabolism, uric acid synthesis and/or renal re absorption, controlling the levels of uric acid in a wide range of values.

In conclusion, despite the great advance of the classic mapping techniques, the isolation of genes involved in speciation and adaptation has been possible only for a few specimens of interest, allowing just a first approach to the genomics of groups of interest. This is the reason why the implementation of new methodologies that hasten the process of generation of highly polymorphic markers and that yield a finer description of the genome and higher resolution maps is of vital importance.

References

Avise, J. 2004. Molecular Markers, Natural History and Evolution. Second Edition. Sinauer Associates, Inc. Publishers. Sunderland, Massachusetts.

Baxter, S., Papa, R., Chamberlain, N., , Humphray, S., Joron, M., , ffrench-Constant, R., McMillan, W.O. & Jiggins, C. 2008. Parallel evolution in the genetic basis of Müllerian mimicry in Heliconius butterflies. Genetics

Collard, B., Jahufer, M., Brouwer, B. & Pang, E. 2005. An Introduction to Markers, Quantitative Trait Loci (QTL) Mapping and Marker-Assisted Selection for Crop Improvement: The Basic Concepts. Euphytica 142: 169 – 196

Colosimo, P., Hosemann, K., Balabhadra, S., Villarreal G., Dickson, M., Grimwood, J., Schmutz, J., Myers, J., Schluter, D., Kingsley, D. 2005. Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. Science 307, 1928 (2005)

Coyne, J. A. & Orr, H. A. 2004 Speciation. Sinauer Associates. Sunderland, Massachussets.

Ellegren, H. 2008. Sequencing goes 454 and takes large-scale genomics into the wild. Molecular Ecology. News and Views.

Futuyma, D. 2005. Evolution. Sianuer Associates, Inc. Publishers. Massachussets, U.S.A.. Chp 9: Variation.

Hartl, D. & Jones, E.. 2005. Genetics. Analysis of Genes and Genomes. Sixth Edition. Jones and Bartlett Publishers. Massachussets, U.S.A. Chp2: DNA structure and Manipulation..

Hui, B. 1998. Statistical Genomics: Linkage Mapping and QTL Analysis. CRC press. USA. Pp 18 – 82.

Iañez, E. 1997. Introducción a los proyectos genoma: Estrategias del PGH. Un iversidad de Granada, España.

Kapan, D., Flanagan, N., Tobler, A., Papa, R., Reed, R., Acevedo, J., Ramirez, M., Martínez, L., Maldonado, K., Ritschoff, C.,Heckel, D. & McMillan, O. 2006. Localiozation of Mullerian mimicry genes on a dense linkage map of *Heliconius erato*. Genetics 173: 735–757

Kronforst, M., Young, L., Kapan, D., McNeely, C., O'Neill, R., & Gilbert, L. 2006. Linkage of butterflie mate preferenceand wing color preferencecue at genomic location of *wingless*.PNAS 103 (17): 6575- 6580

Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln & L. Newburg, 1987. Mapmaker an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174–181

Miller, M., Dunham, J., Amores, A., Cresko, W. & Johnson, E. 2007. Rapad and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Research. www.genome.org.

Orr, H., Masly, J., Presgraves, D. Speciation genes. Genetics & Development 2004, 14:675–679.

Papa, R., Morrison, C., Walters, J., Counterman, B., Rui, C, Halder, G., Roberts, L., Kapan, D., Jiggins, C., Reed, R., & McMillan, W. 2008. Highly conserved gene order and númerous novel repetitive elements in genomic regions linked to wing pattern variation in Heliconius butterflies. Genomics

Peichel, C., Nereng, K., Ohgi, K., Cole, B., Colosimo, P., Buerkle, C., Schluter, D. & Kinsley, D. 2001. The genetic architecture of divergence between threespine stickleback species. Nature 414: 901-905

Thompson, M., Robert L. Nussbaum, Roderick R. McInnes, James Scott Thompson, Huntington F. Willard, Joaquín Salvador Peral, Marta Sedó Fernández. Genética en Medicina: Thompson& Thompsom. Cap8: Mapeo génico y el proyecto genoma humano.

Vera, C., Wheat, C., Fescemyer, H.,Frilander, M., Crawford, D., Hansk, I. & Marden, J. 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Molecular Ecology 17: 1636-1647.

Wang, B. & Porter A. 2004. An AFLP-Based Interspecific Linkage Map of Sympatric, Hybridizing Colias Butterflies. Genetics Society of America 168: 215–225

Yasukochi, Y. 1998. A Dense Genetic Map of the Silkworm, Bombyx mori, Covering All Chromosomes Based in 1018 Molecular Markers. Genetics 150:153 – 1525.

Yasukochi, Y., Ashakumary, L., Baba, K., Yoshido, A. & Sahara, K. 2006. A Second-Generation Integrated Map of the Silkworm Reveals Synteny and Coserved Gene Order Between Lepidopteran Insects. Genetics 173: 1319 - 1328

Zhong, D.,Pai, A.& Yan, G. 2004. AFLP-Based Genetic Linkage Map for the Red Flour Beetle (Tribolium castaneum). Journal of Heredity 95 (1): 53-61

Boxes and Figures

Box 1. Classic Techniques

AFLP: generation of numerous segregant markers from a single reaction of PCR. It is possible to make multiple loci analysis (Hui, 1998; Collard *et al*., 2005; Avise, 2004)
SCNL: useful for mapping genes of interest and as anchor loci for comparative mapping studies through different taxonomic levels. (Avise, 2004)
RFLP: Involves the restriction of DNA with one or more endonucleases, generating a digestión profile tahta may be the result of base substitutions in the cleavage site, DNA insertions or deletions or sequences re arrangements. These are widely used in mapping studies due to their high reproducibility, locus specificity, codominant nature, and neutrality (Avise, 2004; Hui, 1998)
RAPD: This technique involves the use of short PCR primers (around 10bp) to amplify random unknown sequences, allowing the identification of DNA polymorphisms. In addition, it enables the easy and fast detection of potential molecular markers of multiple loci (Avise, 2004; Hui, 1998)
STR: Generates a great quantity of polymorphic markers and they can be used in comparative mapping studies since they can be transferred through populations (Collard *et al*., 2005; Avise, 2004; Hui, 1998)

Box 2. Mapping in Insecta

***Tribolium castaneum*:** Important pest of stored grains. Zong *et al*. (2004) built a linkage map based in AFLPs: The final map was composed of 269 AFLP and 18 RAPD markers, with a resolution of 2cM and around 350kb/cM

***Bombyx mori*:** Yasukochi (1998) built the first linkage map for the silk moth. Afterwards, Yasukochi *et al*. (2006) built a second-generation map with BACs libraries, which allowed them to find synteny and conserved order of genes in four linkage groups with *Heliconius melpomene*, a neotropical butterfly filogenetically distant from it.

Colias eurytheme y Colias phylodice: Economic pests of alfalfa and clover crops. Wang & Porter (2004) generated a linkage map based on AFLPs with 51 linkage groups, a total of 2541,7 cM.

***Heliconius erato*:** Kapan *et al*. (2006) constructed a second-generation map based on AFLP, SCNL, and microsatellites. 20 linkage groups and the Z chromosome were found. They also located two color pattern loci: *D*, which codifies for the red or orange elements and Sd, which controls the melanism of the forewing

***Heliconius melpomene, Heliconius erato y Heliconius numata*:** Baxter (2008**)** made an approach through positional cloning in *H. melpomene* that led to the identification of markers linked to a color pattern locus *Yb*, which controls the yellow band on the hindwing. These were mapped in *H.*erato and *H.numata*, and it was established that they were tightly linked to the *H.erato* locus *Cr* that has similar effects to those of *Yb*, and to the *H.numata* locus *P*, which controls all the phenotypic polymorphism of the wing. These results point out that one single genetic locus is controlling the convergent and divergent phenotypes in different lineages.

Preference Mapping: Kronforst *et al*. (2006) through a QTL and AFLP analysis in *Heliconius* butterflies, located a preference for color locus that coincides with the locus that controls the forewing color, which itself is perfectly linked to the pattern candidate gene *wingless*

Box3. Genes de Adaptación y Especiación

- ***Drosophila*:** In the fruit fly it has been possible to isolate genes that are involved in hybrid sterility and inviability (Orr *et al*., 2004):
  ***OdsH (Odisseus site homeobox)*:** hybrid sterility in hybrid males of *D.simulans* and *D.mauritiana*
  Hmr (Hybrid male rescue): Detention in the development in the larvae-pupae transition.
  Nup96 (Nucleoporin 96): the allele of *Nup96* in *D.simulans* in incompatible with an unknown gene in the X chromosome of *D.melanogaster*

- ***Xiphophorus*** (Orr *et al*., 2004)
  ***Xmrk2*:** Codifies for a tyrosine-kinase receptor that is over-expressed in *Xiphophorus* hybrids causing lethal tumorogenesis and that is in association with a pigmented pattern locus.

- ***Coereba flaveola*** *(*Theron *et al*., 2001): In the bananaquits, a gene responsible for the plumage polymorphism of two ecologically isolated populations of the same species was located.
  ***MC1R*:** Key regulator of the melanin synthesis in the feather melanocytes that causes the difference of the phenotypes produced by a simple genetic change from glutamate to lysine.

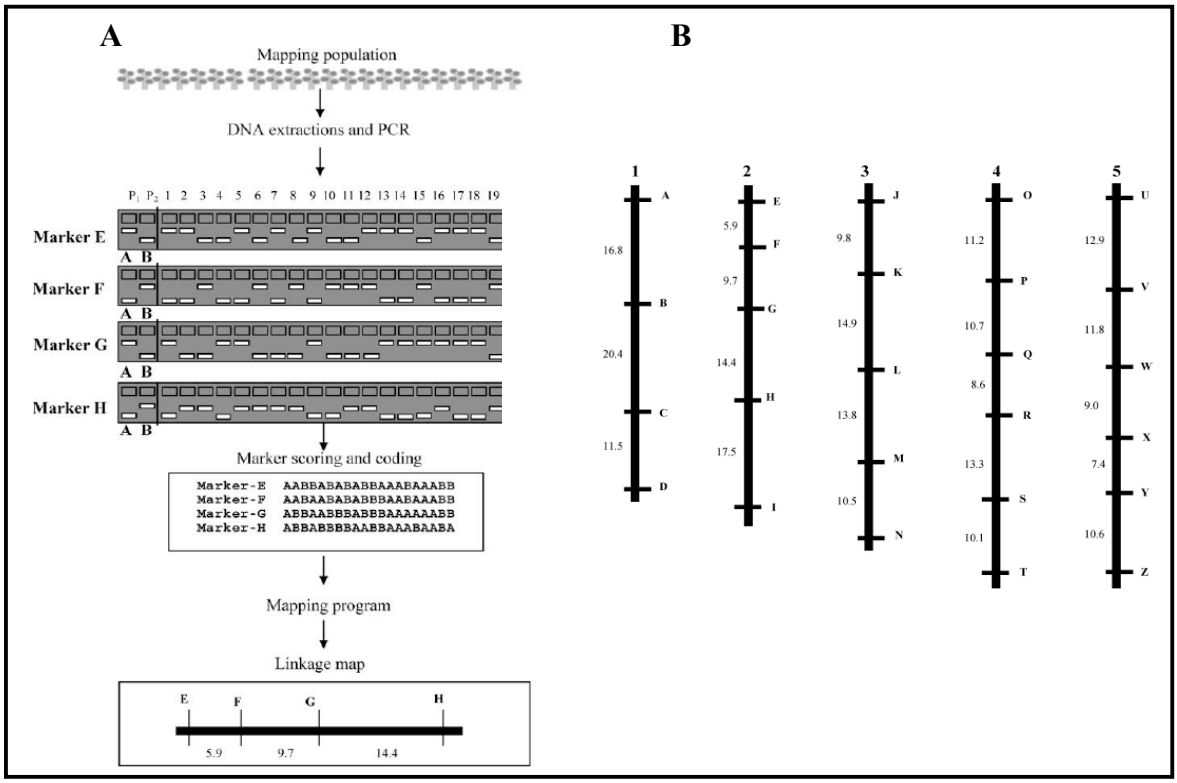

Figure 1. Linkage-map construction and output. (A) The first parent (P_1) is presented as "A" whereas the second parent (P_2) is presented as "B". The codification of the markers data varied depending on the population used. This linkage map was constructed with MapManager using Haldane's mapping function. (B) Hypothetical linkage map of five chromosomes (represented by five linkage groups) and 26 markers. Ideally, a linkage map should have markers uniformly distributed for a subsequent QTL analysis. In addition, it should also contain anchor loci present in other linkage maps in order to conduct comparative analyses of genomic regions between different maps. Taken from Collard *et al*. (2005)

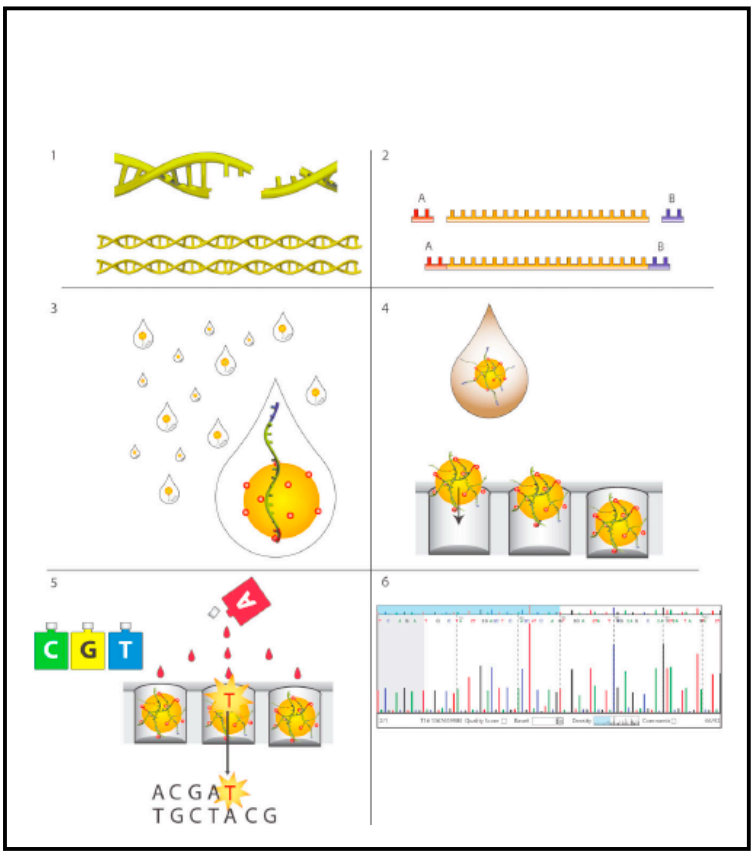

Figure 2. A schematic illustration of several steps in 454-sequencing. (1) Short amplicons, fragmented DNA or cDNA is used as starting template for further processes. (2) 5′- and 3′-end specific adapters are ligated to single-strand fragments, creating a library for PCR amplification. (3) Fragments are immobilized through a biotin tag on one of the adaptors that bind to streptavidin-coated beads. Each bead will come to carry just one fragment. Beads are then emulsified in a water-in-oil mixture. (4) Each drop of oil contains the necessary ingredients for PCR and thereby forms a microreactor for amplification. Massively parallel amplification is carried out in the emulsion. Beads, with amplified fragments bound to them, are released from oil and are loaded onto a fibre optic chip, a picotiter plate, for sequencing. Only one bead will fit in each ~ 44 µm well. (5) Pyrosequencing takes place by a sequential flow of sequencing reagents across the plate. When a complementary nucleotide is added to a particular template in an extension reaction, a light signal is generated. (6) The final result is a pyrogram in which the height of each signal is proportional to the number of adjacent nucleotides that are identical. In each cycle of the sequential addition of the four different nucleotides, no signal is seen when noncomplementary nucleotides are added. Taken from Collard *et al*. (2005)

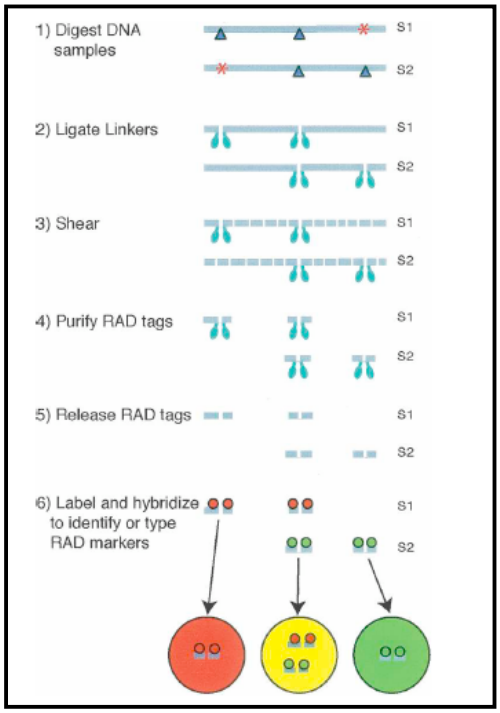

Figure 3. Restriction site associated DNA (RAD) markers can be identified by detecting differential hybridization patterns of RAD tags on a microarray. Genomic DNA samples S1 and S2 contain the recognition sequence for various restriction enzymes at locations throughout the genome. Dark blue triangles represent restriction sites of a particular enzyme. Some of these restriction sites are only present in one sample because of polymorphisms that disrupt the recognition sequence (red asterisks). Both samples are separately digested with a particular restriction enzyme and then ligated to biotinylated linkers (light blue ellipses). The DNA is randomly sheared leaving only the fragments that were directly flanking a restriction site attached to biotin linkers. These fragments are purified using streptavidin beads and released by digestion at the original restriction site. Loci containing polymorphisms, such as the *left* locus of S2 or the *right* locus of S1, will not contain tags for that locus in the purified RAD-tag sample, thus resulting in differential hybridization patterns of RAD tags on a microarray. Taken from Miller *et al*. (2007).