9. Genetic and Physical Maps

Karen Marshall

Learning objectives

By the end of this topic, you should have:

- an understanding of the terms genetic and physical maps, and the units used to measure distances on these maps
- the ability to construct a simple genetic map from three-point testcross data
- the ability to relate observed recombination frequencies to map distances through mapping functions
- an understanding of the concept that multi-locus linkage maps from complex pedigree can be constructed by the maximum likelihood approach
- an understanding of the different types of physical maps and how they are constructed
- conceptualised how genetic and physical maps can be linked together
- an overview of the current genome sequencing status for a number of livestock species

Key terms and concepts

Genetic map, recombination frequency, centiMorgans (cM), three-point testcross, interference, mapping functions, multi-locus linkage maps, physical maps, cytogenetic maps, radiation hybrid (RH) maps, STS maps, sequence maps, whole genome shotgun sequencing, clone contig or hierarchical shotgun sequencing, linkage of genetic and physical maps, variable relationship between cM and megabase pairs (Mb).

Introduction to the topic

Genetic and physical maps are used in QTL mapping and gene discovery experiments. These maps show the location of genetic markers, genes, and other features (such as repeat sequences) on the genome. This topic overviews the features of these maps and how they are constructed.

9.1 Overview

A genetic map is a set of genetic markers that have been assigned to particular chromosomal locations based on linkage analysis. The distance between genetic markers is measured in Morgans or centiMorgans (cM), calculated from observed recombination frequencies. The genetic map describes the behaviour of genetic markers as they are transmitted from one generation to another, i.e. whether markers tend to segregate together during meiosis because the markers are located close together on the same chromosome.

A physical map of an organism describes the actual structure of the genetic material, not its behaviour. In its most detailed form the physical map is the DNA sequence of each chromosome. The physical map is measured in nucleotides or base pairs, and typically presented in units of 10^6 base pairs as megabase pairs (Mb). Genetic and physical maps are typically linked together.

The most detailed maps are available for the human genome. Under the NCBI map viewer, http://www.ncbi.nlm.nih.gov/mapview/, close to 50 different human maps can be selected. These include numerous sequence maps (showing for example repeat sequences and gene sequences), other physical maps such as cytogenetic and radiation hybrid maps, and genetic linkage maps. A partial map of human chromosome 1 is shown in Figure 9.1.

©2009 The Australian Wool Education Trust licensee for educational activities University of New England

Figure 9.1 Partial map of Human chromosome 1[1].
Source: National Library of Medicine (2004).

[1] Displayed maps from left to right are 1) **Ideogram:** ideogram of a cytogenetic map, 2) **NCBI_RH**: NCBI integrated radiation hybrid map, 3) **Repeats:** showing position of repetitive elements from sequence data, 4) **Genethon:** a genetic linkage map comprised of microsatellites 5) **DeCode:** another genetic linkage map 6) **Genes_seq**: genes that have been annotated including known and putative genes, different colours are used to indicate difference levels of evidence for the gene.

©2009 The Australian Wool Education Trust licensee for educational activities University of New England

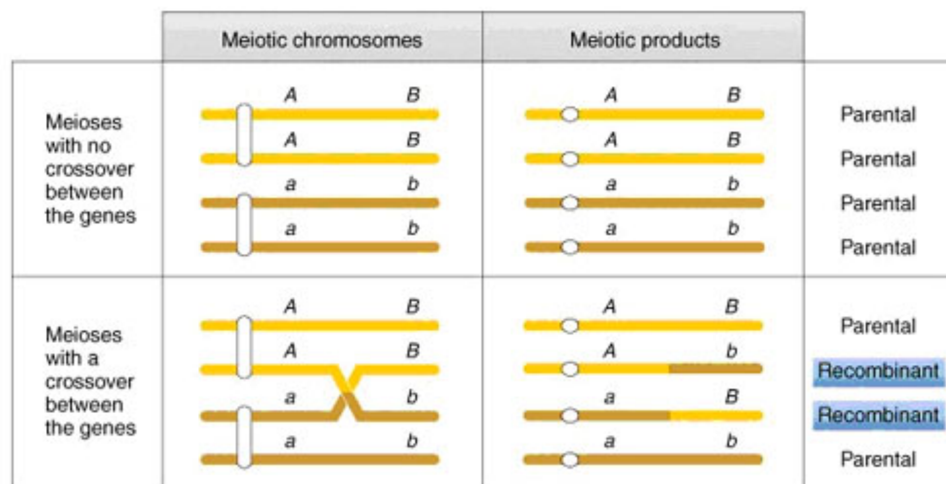
9.2 Genetic maps

Map distance relates to recombination frequency

Genetic maps (also termed linkage maps) are constructed from observations of recombination frequency between genetic markers. Recall that two loci on different chromosomes show independent assortment and are described as unlinked. An AaBb individual (where *A* and *B* are unlinked loci with alleles A,a and B,b respectively) will produce AB, Ab, aB, and ab gametes in equal frequencies. If, however, the two loci are linked independent assortment does not occur and the four possible gametes are produced in frequencies dependant on the recombination frequency (r) between the two loci. Each of the parental gametes will be produced at a frequency of $(1-r)/2$, and each of the re-combinant gametes at a frequency of $r/2$ (Figure 9.2).

Under a simple mapping function, which assumes map distance is directly proportional to the number of observed recombinants, one recombinant gamete out of 100 gametes corresponds to a map distance of 1cM. (We see later that assumptions of this simple mapping function do not always hold).

Figure 9.2 Meiotic recombination. Source: Krempels (2004).



Markers used for genetic maps

Any type of genetic marker can be used to construct a genetic map. The most common type is molecular markers (Topic 8). Other types of markers include mutations resulting in a defined phenotype and protein polymorphisms.

A simple genetic mapping example

Consider a classic mapping example of a three-point testcross in *Drosophila*, as analysed by Sturtevant (a student of Morgan's). Genes for an eye, wing and body trait, known to be located on the X chromosome, were under investigation. Recessive alleles are notated 'w', 'min' and 'y', for white eye, miniature wing and yellow body, respectively.

Data was produced from the following testcross:
F1 females [w/+ . min/+ . y/+] x tester male [w/⇀ . min/⇀. y/⇀].
Note that + refers to the corresponding wild-type allele, ⇀ refers to the Y chromosome, and that the order of which the loci are written (*eye, wing, body*) is currently arbitrary. Further prove to yourself that the phenotypes of both the male and female progeny from this testcross directly reflect the gamete type passed on by the F1 female.

©2009 The Australian Wool Education Trust licensee for educational activities University of New England

Table 9.1 summarises the number of testcross progeny receiving the different gamete types from the F1 female. A number of deductions can be made from this data. Firstly, the parental gametes can be identified as <u>+ + +</u> and <u>w min y</u>, as these are the most represented of the 8 gamete types. Thus, it can be deduced that the parents of the F1 female were [w/w . min/min . y/y] x [+/- . +/-. +/-]. Secondly, the recombinant gamete types can be identified. For example, <u>+ min +</u> is recombinant for *eye* and *wing* (as <u>+ min</u> is not a parental gamete type) and *wing* and *body* (similarly, as <u>min +</u> is not a parental gamete type). Thirdly, the relative position and distance between the three loci can be determined from the recombination frequencies. The recombination frequency (*r*) is simply calculated as the number of recombinant gametes / the total number of gametes. Under the simple mapping function multiplication of *r* by 100 gives genetic distance in cM.

Figure 9.3 shows the resultant genetic map. The greatest recombination frequency (0.3383) is for *wing-body*, which are thus the two 'outer' loci. The *eye* loci lies between the two, close to the *body* loci as indicated by the small recombination fraction (0.0131%) for *eye-body* (Figure 9.3). Note that the recombination frequencies do not quite add up: 0.3261 (*wing-eye*) + 0.0131 (*eye-body*) equals 0.3392 and not 0.3383 (*wing-body*). This small discrepancy is because double-crossovers have not been accounted for. Rewriting the parental gamete in the correct order of <u>y w min</u>, it is evident that two cross-over events are required for <u>y + min</u> and <u>+ w +</u>. Counting these recombinant classes twice gives the recombination fraction for *wing-body* of (401+317+16+12+**2**) / 2205 = 0.3392, the sum of the two component values.

Table 9.1 Drosophila three-point testcross data. Source: Johnson, (date unknown).

Gamete (eye wing body)	**Number progeny**	**Recombinant for**		
		eye wing	***eye body***	***wing body***
+ + +	758			
w min y	700			
+ min +	401	401		401
w + y	317	317		317
w min +	16		16	16
+ + y	12		12	12
w + +	1	1	1	
+ min y	0	0	0	
Total	***2205***	***719***	***29***	***746***
r		***0.3261***	***0.0131***	***0.3383***

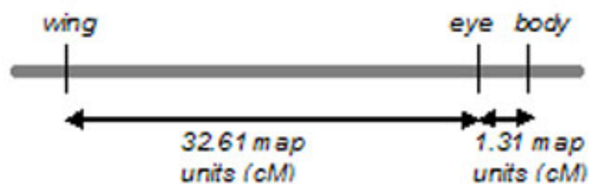


Figure 9.3 Map deduced from the drosophila three-point testcross data presented in Table 9.1.

©2009 The Australian Wool Education Trust licensee for educational activities University of New England

Interference

Using recombination frequency as a measure of genetic distance implies that loci twice as far apart show twice as many cross-overs, i.e. a linear relationship. However it is known that a crossover in one region can affect the likelihood of a crossover in an adjacent region. This is referred to as interference. If all crossover events occurred independently, the frequency of double crossovers should equal the product of the frequency of single crossovers. In the above example the frequency of double crossovers was 0.045% (1/ 2205), and the product of the single crossovers 1.31% x 32.61 % = 0.42%. Clearly the data does not fit a no-interference model. The level of interference is actually 1 – (observed frequency of double recombinants / expected frequency of double recombinants) = 1 – (0.045/0.42) = 0.89 or 89%. In this case it is a relatively high value, due to the small distance between the eye and body loci. Positive interference is often observed in eukaryotes, the level of interference decreasing as loci become further apart. In *drosophila*, it has been observed that double crossovers do not occur in regions of 10-12cM or smaller. Interference is likely due to physical constraints preventing the formation of closely aligned chiasmata (Klug and Cummings, 2005).

Unobservable crossovers and mapping functions

An odd-number of cross-over events is observed, whereas an even number is not (as an activity, draw a diagram to prove this to yourself). Thus the observed number of recombination events is likely lower than the real number. To account for this, mapping functions are used. Mapping functions predict map distance (m) from the observed recombination frequency (r).

Two mapping functions in common use are those of Haldane (1919) and Kosambi (1944). Haldane's map function assumes crossovers occur both randomly and independently:

$$m = -\frac{\ln(1-2r)}{2}$$

whereas Kosambi's mapping function allows for modest interference (Lynch and Walsh, 1998):

$$m = \frac{1}{4}\ln\frac{(1+2r)}{(1-2r)}.$$

Figure 9.4 shows the relationship between r and a simple mapping function where $m = r$, Haldane's map distance and Kosambi's map distance. Note that for small r the map distances are similar, whereas for r values approaching 0.5 the Kosambi and particularly Haldane map distances are much greater than r.

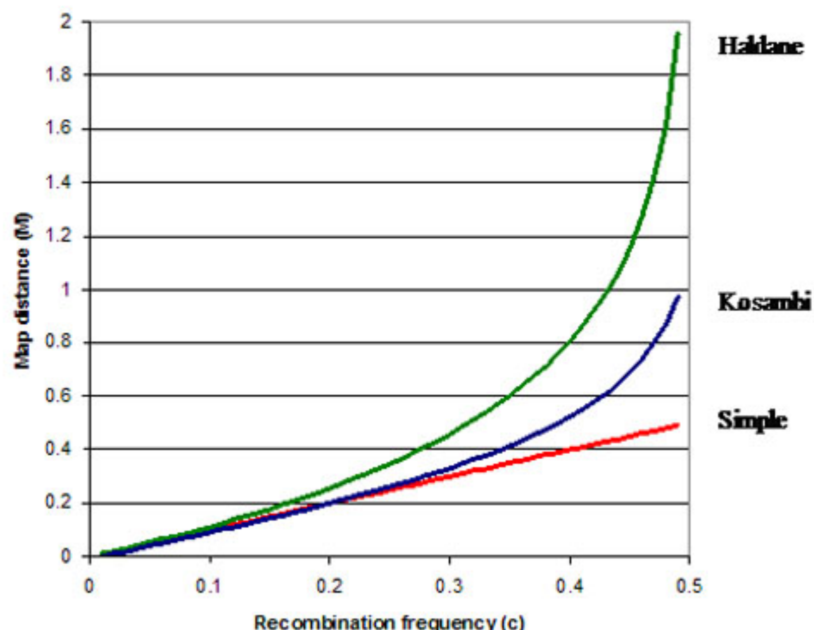


Figure 9.4 Comparison of Haldane, Kosambi and simple (m=r) mapping functions. Source: Adapted from Weller (2001).

©2009 The Australian Wool Education Trust licensee for educational activities University of New England

Maps from complex pedigrees using a maximum likelihood approach

In practice, genetic linkage maps are often constructed for complex pedigrees using a maximum likelihood (ML) approach. The ML approach is used as it allows for an unknown linkage phase in the parents. In designed experiments (such as the three-point testcross described above) the linkage phase is obvious as the number of non-recombinant progeny is much greater than the number of recombinant progeny. However in complex pedigrees with smaller family sizes this may not be the case. The process of ML determines the *most likely* phase from available data and thus the most likely recombination frequency. (Gene522 students should refer to Van der Werf et al. (2003) listed in the readings section for a small mapping example using the maximum likelihood approach).

Often many loci are mapped simultaneously, a procedure referred to as multilocus linkage analysis. A number of software packages are available to perform such analysis, such as cri-map (http://compgen.rutgers.edu/ multimap/ crimap/).

9.3 Physical maps

Types

The main categories of physical maps are cytogenetic maps, radiation hybrid (RH) maps, sequence tag site (STS) maps and sequence maps. Each of these will be briefly described below.

Cytogenetic maps are based on the banding patterns of stained chromosomes under light microscopy (Figure 9.5). These are the lowest resolution type of map: the number of base pairs within a band can only be estimated.

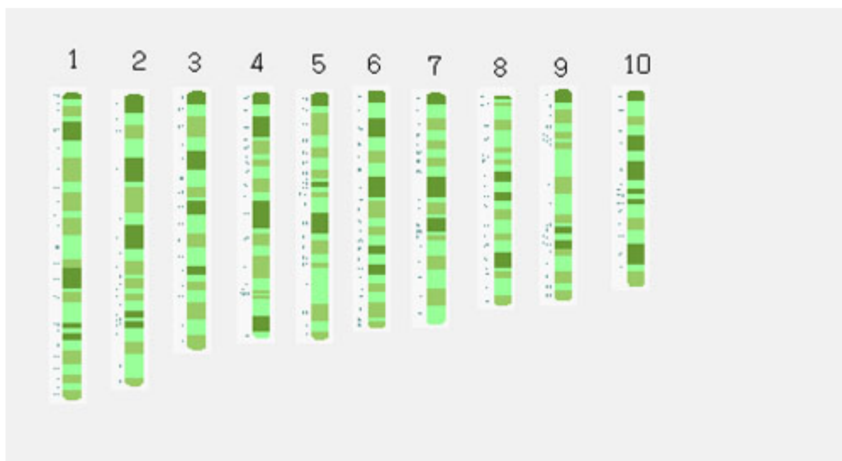


Figure 9.5 Mouse cytogenetic map for chromosomes 1 to 10.
Source: Animal Genome Research Program (2004).

Radiation hybrid (RH) mapping is similar to genetic mapping, except that RH maps estimate the distance between two marker loci from the number of breaks induced by radiation. The procedure to construct a RH map is as follows. A culture of diploid cells from the species to be mapped is irradiated, resulting in fragmentation of the chromosomes and death of the cells. To allow propagation of the fragmented chromosomes, the irradiated cells are fused to a recipient cell line, resulting in donor DNA fragments contained within recipient chromosomes (Figure 9.6). DNA from the clones is then tested for the presence of markers or genes. Closely linked markers are likely to be retained or lost on a single fragment, whereas distant markers are more likely to be retained or lost independently. The actual distance and order between markers is calculated using likelihoods, the distance between two loci on a RH map being proportional to the physical distance between the loci. The average resolution between markers is often 2 – 10 Mb. Varying the radiation dose varies the fragment size, thus allowing the resolution to be manipulated.

©2009 The Australian Wool Education Trust licensee for educational activities University of New England

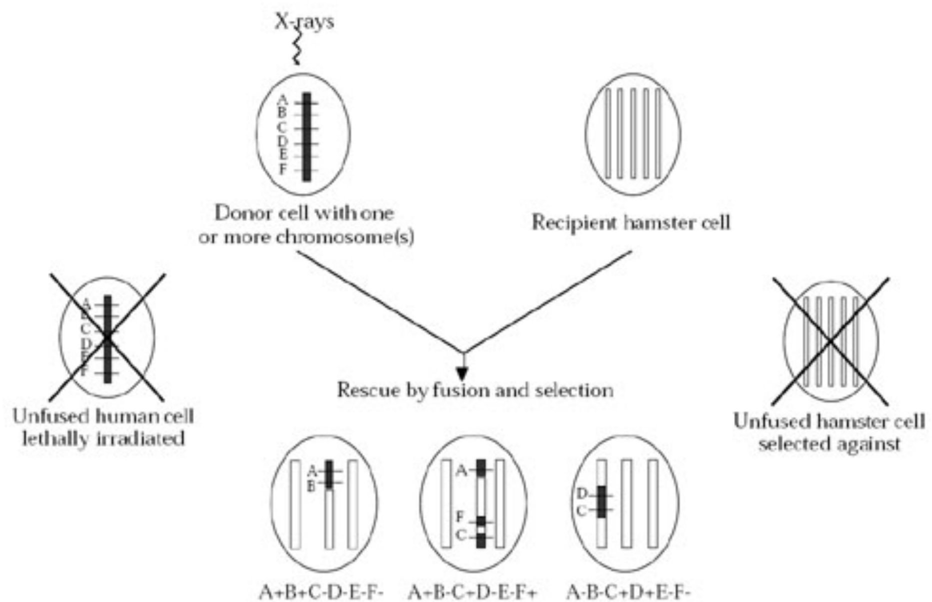


Figure 9.6 Radiation hybrid construction. Source: Matise (1999).

A **genome sequence** represents the ultimate physical map. The basic steps in creating a sequence map (of which there are many variations) are as follows. The genome is broken into smaller fragments which are cloned, to form a gene library. These clones are then sequenced, and overlapping sequences assembled into contigs, until a consensus sequence of the genome is reached. Note that to ensure the consensus sequence does not contain genotyping errors, each base pair is usually read by 5-10 independent sequence reads (referred to as e.g. 5 fold coverage).

"Whole genome shotgun" sequencing refers to a strategy where clones from a genome library are randomly selected and sequenced, and the sequence then assembled (Figure 9.7). In this strategy there is no prior information on where the clones map to the genome. This approach has most commonly been used for prokaryotic (bacterial) genomes, which are relatively small and do not contain repetitive sequences that interfere with sequence assembly. In eukaryotes a derivation of the approach (using pair-end reads) has been useful in providing draft sequence and ordering contigs into larger units known as scaffolds.

An approach more suitable for eukaryotic genome sequencing is the "clone contig" or "hierarchical shotgun" approach. In this case the clones are first ordered into a contig, a minimal set of representative clones identified, and these clones then sequenced by the shotgun approach (Figure 9.8). The clones used for contig formation generally contain large inserts, and are often generated using bacterial artificial chromosomes (BACs) as a cloning vector with inserts of 100-200kb each. The clones can be ordered into the contig based on overlapping STS markers (Figure 9.9, STS markers are defined below), or by the more laborious chromosome walking approach. A contig based on overlapping STS markers is the basis of a **STS map**.

It is interesting to note that in sequencing the human genome both approaches were employed: the International Human Genome Sequencing Consortium used the hierarchical shotgun approach and the private biotechnology company Celera Genomics used the whole genome shotgun approach.

Note: students not familiar with the term gene library and chromosome walking should see the appropriate reference under readings.

©2009 The Australian Wool Education Trust licensee for educational activities University of New England

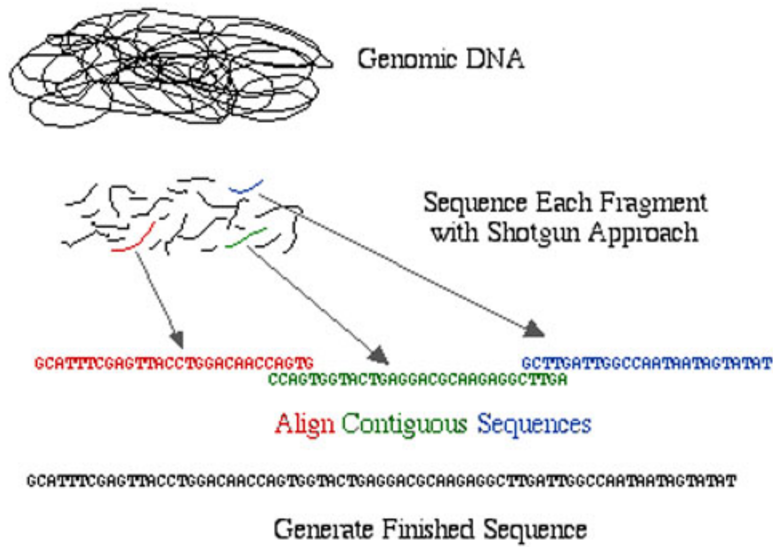


Figure 9.7 Whole genome shotgun sequencing. Source: Campbell (2004).

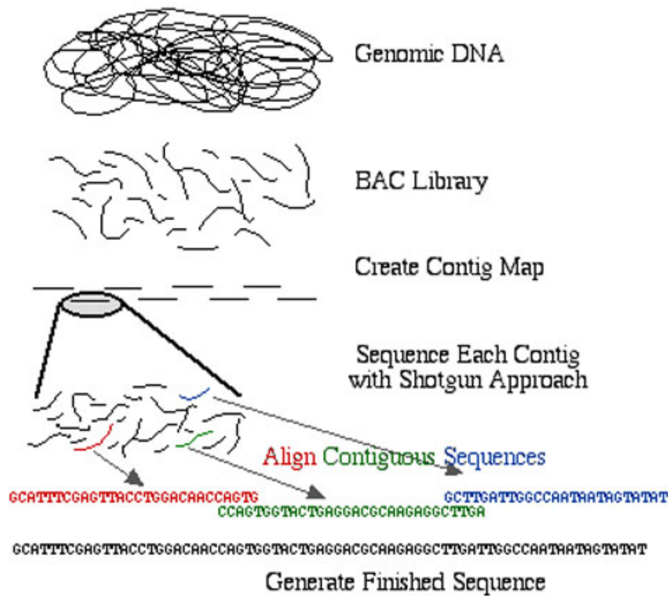


Figure 9.8 The hierarchical shotgun sequence approach. Source: Campbell (2004).

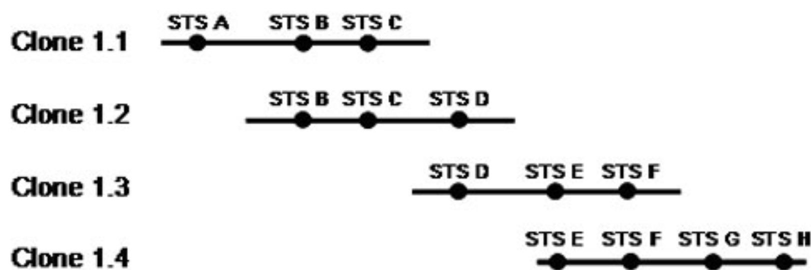


Figure 9.9 Construction of a clone contig, using markers such as STS to order clones.

©2009 The Australian Wool Education Trust licensee for educational activities University of New England

Markers used for physical maps

Markers used in physical map construction, such as in RH mapping or to order clones to construct a clone contig, are often **sequence tag sites** (STSs). STSs are unique sequences of 100-500 bp in length, generally amplifiable by PCR. The uniqueness of these sequences is important, as non-unique sequence would lead to ambiguous mapping results. STSs can be random genomic sequences, identified from data deposited into sequencing databases. Alternatively many STSs are derived from **expressed sequence tag** (EST) sequences, found as follows. mRNA sequence (i.e. the expressed sequence) is back-transcribed to complimentary DNA (cDNA) which is partially sequenced (Figure 9.10) - this partial sequence represents the EST. An EST can be used as an STS if it is derived from a unique gene.

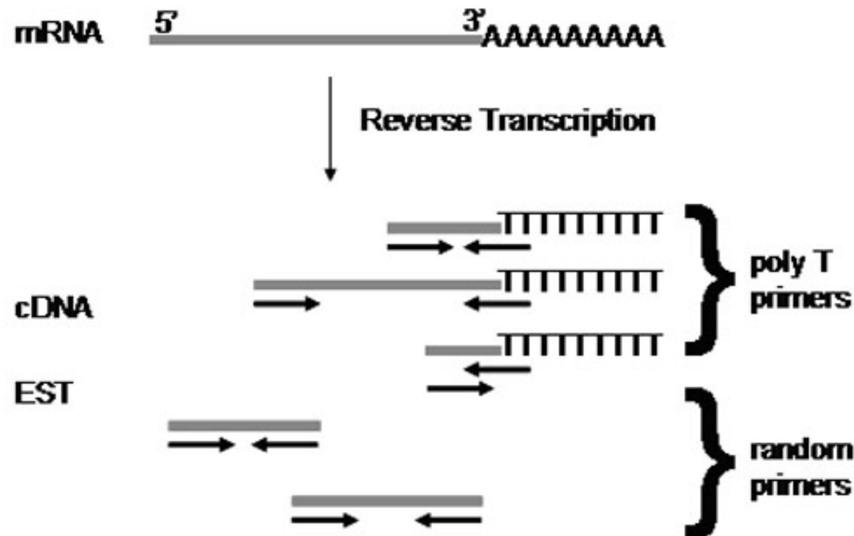


Figure 9.10 Expressed sequence tags (ESTs). The EST sequence is a partial sequence of the cDNA, as represented by arrows.

9.4 Linkage of genetic and physical maps

Genetic maps require segregating genetic polymorphisms and thus mainly contain highly polymorphic markers such as microsatellites. In contrast, physical maps are created by directly assigning a gene or marker to a chromosomal location. As a genetic polymorphism is not required for physical maps, these maps usually contain a higher number of expressed genes. To maximise the utility of both types of maps, i.e. genetic and physical, they are usually linked together.

Linkage of genetic and physical maps can be achieved in a number of ways. For example, molecular markers used on a genetic map can be located on the genomic clones used in constructing the physical map or matched to the actual genomic sequence.

Note that there is no universal relationship between genetic distance (cM) and physical distance (Mb), although ‘rules of thumb’ are often applied. The relationship varies, dependant on the species, the chromosomal region (as crossovers are often suppressed at centromeres and telomeres), and in some cases the sex of individuals (for example, crossovers do not occur in male drosophila).

The following is taken from a paper titled “Comparison of human genetic and sequence-based physical maps” by Yu et al. (2001).

©2009 The Australian Wool Education Trust licensee for educational activities University of New England

"*The completion of the draft human genomic sequence provides us with the best opportunity yet to compare the genetic and physical maps. Here we describe our estimates of female, male and sex-average recombination rates for about 60% of the genome. Recombination rates varied greatly along each chromosome, from 0 to at least 9 centiMorgans per megabase (cM Mb^{-1}). Among several sequence and marker parameters tested, only relative marker position along the metacentric chromosomes in males correlated strongly with recombination rate. We identified several chromosomal regions up to 6 Mb in length with particularly low (deserts) or high (jungles) recombination rates.*"

9.5 Current state of maps for various species

The status of species maps is constantly changing. At present (June 2007), NCBI Map Viewer (a major public database of map information: http://www.ncbi.nlm.nih.gov/mapview/), displays maps for the following vertebrates: human, chimpanzee, mouse, rat, chicken, cow, pig, sheep, dog, cat, zebrafish, rhesus macaque, and an opossum. A short statement on the current status in relation to genome sequencing for several of these follows.

Human: The human genome has been fully sequenced and annotated. The Human Genome Project (HGP) was a major international collaboration which sequenced the near 3 billion base pairs of the human genome and identified approximately 30,000 human genes. The project began in the 1990's, a draft sequence was published in February 2001, and the final sequence in April 2003 (The International Human Genome Mapping Consortium, 2001; The Celera Genomics Sequencing Team, 2001; Collins et al 2003a; Collins et al 2003b; Frazier et al 2003; Carroll 2003).

Updated releases of the sequence assembly and annotation continue to become available. The latest is 'Build 36, version 2', released September 2006. View **http://www.ncbi.nlm.nih.gov/mapview/stats/BuildStats.cgi?taxid=9606&build=36&ver=2** for summary statistics of this build. Note that the number of genes identified differs slightly depending on the methodology used for assembly and annotation. For example, the *reference assembly* gives 28961 placed genes (and 239 not placed), whereas the *Celera assembly* gives 26245 placed genes (and 633 not placed).

Following on from the HGP a HapMap project was initiated, which aimed to identify human DNA variations (http://www.hapmap.org/). A haplotype map of the human genome was published in Nature in 2005, with over 1 million SNPs reported (The International HapMap Consortium, 2005).

Chicken: Sequencing is essentially complete for a Red Jungle Fowl Inbred line (considered to be the ancestor of all domestic chickens) at 7 fold coverage. The first draft of the assembled and annotated sequence was placed in public databases in March 2004. The current build is version 2.1, of May 2006. The number of placed and unplaced genes is 15945 and 3538, respectively.

Cow: The bovine genome has been sequence at a 7 fold coverage, using DNA from an inbred Hereford cow. The first draft of the assembled sequence (for a lower fold coverage) was released in October 2004. The current build is version 3.1, of January 2007. The number of placed and unplaced genes is 23119 and 5487, respectively.

Pig: The pig genome is currently being sequenced (see http://piggenome.org/ for an update).

Sheep: The sheep genome is currently being sequenced (see http://www.sheephapmap.org/), via the efforts of the International Sheep Genomics Consortium, of which Australia (via organisations including CSIRO, AWI and MLA) is a partner.

Dog: Sequencing is essentially complete for an inbred Female Boxer at 7.6 fold coverage. The first draft was released July 2004. The current build is version 2.1, of September 2005. The number of placed and unplaced genes is 29528 and 379, respectively.

©2009 The Australian Wool Education Trust licensee for educational activities University of New England

Readings

1. Maddox, J.F., Davies, K.P., Crawford, A.M, Hulme, D.J., Vaiman, D., Cribiu, E.P., Freking, B.A., Beh, K.J., Cockett, N.E., Kang, N., Riffkin, C.D., Drinkwater, R., Moore, S.S., Dodds, K.G., Lumsden, J.M., van Stijn, T.C., Phua, S.H., Adelson, D.L., Burkin, H.R., Broom, J.E., Buitkamp, J., Cambridge, L., Cushwa, W.T., Gerard, E., Galloway, S.M., Harrison, B., Hawken, R.J., Hiendleder, S., Henry, H.M., Medrano, J.F., Paterson, K.A., Schibler, L., Stone, R.T. and van Hest, B. 2001, 'An enhanced linkage map of the sheep genome comprising more than 1000 loci', *Genome Research*, vol. 11(7), pp. 1275-1289.
2. Weller, J. I. 2001, 'Genetic mapping functions', in *Quantitative trait loci analysis in animals*, CAB International, New York, pp. 9-12.

Gene522 students:

Genetic mapping using the ML approach

1. Van der Werf, J., Kinghorn, B., Hayes, B. and Goddard, M. 2003, 'Estimation of recombination fraction', 'ML estimation of linkage' and 'testing for linkage: LOD scores', in *QTL mapping for practitioners, from linkage to genes* UNE Animal Breeidng Summer Course pp. 15-20. http://www-personal.une.edu.au/~jvanderw/aabc_materialsp3.htm

For students unfamiliar with the terms gene library and chromosome walking, the following references will assist you but are not provided.

Gene libraries

4. Griffiths, A.F., Wessler, S.R., Suzuki, D.T., Lewontin, R.C. and Gelbart, W.M. 2005, Making genomic and cDNA libraries', in Introduction to genetic analysis 8th Ed, W. H. Freeman and Company, New York, pp. 350-351.

Chromosome walking

5. Griffiths, A.F., Wessler, S.R., Suzuki, D.T., Lewontin, R.C. and Gelbart, W.M. 2005, 'Figure 11-15', in *Introduction to genetic analysis 8th Ed*, W. H. Freeman and Company, New York, pp. 356.

Summary

Genetic and physical maps are tools used in QTL detection and gene mapping. The underlying basis of the two maps differs. Genetic maps are based on observed recombination frequencies between loci, with genetic distance expressed in Morgans. Physical maps are based on the actual structure of the genetic material, with genetic distance expressed in base pairs. Although genetic and physical maps are generally linked, the relationship between genetic and physical distance is not universal. Advances in molecular biology techniques over recent years has resulted in the development of maps for most livestock species, with full genome sequence available or becoming available for several species.

©2009 The Australian Wool Education Trust licensee for educational activities University of New England

Reference

Animal Genome Research Program, 2004, 'Mouse cytogenetic map image', retrieved 1st July 2005 from http://ws4.niai.affrc.go.jp/dbsearch2/mmap/ mmap.html.

Campbell, A.M. 2004, 'Sequencing whole genomes' in Bio309 Genomics, Proteomics and Systems Biology Course, Department of Biology, Davidson College, Davidson, retrieved 1st July 2005 from http://www.bio.davidson.edu /courses/genomics/method/shotgun.html.

Carroll, S.B. 2003, 'Genetics and the Making of *Homo sapiens', Nature,* vol. 422, pp. 849-857.

Celera Genomics Sequencing Team, 2001, 'The sequence of the human genome', *Science,* vol. 291, pp. 1304-1351.

Collins, F.S., Green, E.D., Guttmacher, A.E. and Guyer, M.S. 2003a, 'A Vision for the Future of Genomics Research - A blueprint for the genomic era', *Nature,* vol. 422, pp. 835-847.

Collins, F.S., Morgan, M. and Patrinos, A. 2003b, 'The Human Genome Project: Lessons from Large-Scale Biology', *Science,* vol. 300, pp. 286-290.

Frazier, M.E., Johnson, G.M., Thomassen, D.G., Oliver, C.E. and Patrinos, A. 2003, 'Realizing the Potential of the Genome Revolution: The Genomes to Life Program', *Science,* vol. 300, pp. 290-293.

Griffiths, A.F., Wessler, S.R., Suzuki, D.T., Lewontin, R.C. and Gelbart, W.M. 2005, *Introduction to genetic analysis*, 8th Edn, W. H. Freeman and Company, New York.

Haldane, J.B.S. 1919, 'The combination of linkage values, and the calculation of distance between the loci of linked factors', *Journal of Genetics,* vol. 8, pp. 299-309.

International HapMap consortium, 2005, 'A haplotype map of the human genome', *Nature* vol. 437, pp. 1299-1320.

International Human Genome Mapping Consortium, 2001, 'A physical map of the human genome', *Nature* vol. 409, pp. 934-941.

International Society of Animal Genetics, 2004, *Proceedings of the 29th International Conference on Animal Genetics*, Tokyo, September 2004.

Johnson, G.B. date unknown, 'STURTEVANT - The First Genetic Map: Drosophila X Chromosome', Chapter 4 in *How Scientists Think: Twenty-One Experiments that Have Shaped Our Understanding of Genetics and Molecular Biology,* The McGraw-Hill Companies, New York. Retrieved on 1st July 2005 from http://www.mhhe.com/biosci/genbio/raven6b/graphics/ raven06b/howscientiststhink/04-lab.pdf

Klug, W.S. and Cummings, M.R. 2005, *Essentials of Genetics*. Prentice Hall Inc.

Kosambi, D.D. 1944, 'The estimation of map distances from recombination values', *Ann. Eugen.* vol. 12, pp. 172-175.

Krempels, D. 2004, 'Meiotic recombination' University of Miami Department of Biology, retrieved 1st July 2005 from http://fig.cox.miami.edu/Faculty/Dana/ F05_08.JPG.

Lynch, M. and Walsh, B. 1998, *Genetics and analysis of quantitative traits*, Sinauer Associates, Inc, Sunderland.

Maddox, J.F., Davies, K.P., Crawford, A.M, Hulme, D.J., Vaiman, D., Cribiu, E.P., Freking, B.A., Beh, K.J., Cockett, N.E., Kang, N., Riffkin, C.D., Drinkwater, R., Moore, S.S., Dodds, K.G., Lumsden, J.M., van Stijn, T.C., Phua, S.H., Adelson, D.L., Burkin, H.R., Broom, J.E., Buitkamp, J., Cambridge, L., Cushwa, W.T., Gerard, E., Galloway, S.M., Harrison, B., Hawken, R.J., Hiendleder, S., Henry, H.M., Medrano, J.F., Paterson, K.A., Schibler, L., Stone, R.T. and van Hest, B. 2001, 'An enhanced linkage map of the sheep genome comprising more than 1000 loci', *Genome Research*, vol. 11(7), pp. 1275-1289.

Matise, T.C. 1999, 'Radiation Hybrid Mapping' in 1999 - Current topics in genome analysis, a mini course covering important areas in genome analysis, retrieved 1st July 2005 from http://www.genome.gov /Pages/Hyperion/COURSE99/Pdf/ matise.pdf.

National Library of Medicine, 2004, 'Partial map of human chromosome 1', Retrieved 22nd October 2004 from http://www.ncbi.nlm.nih.gov/mapview/.

Yu, A., Zhao, C., Fan, Y., Jang, W., Mungall, A.J., Deloukas, P., Olsen, A., Doggett, N.A., Ghebranious, N., Broman, K.W. and Weber, J.L. 2001, 'Comparison of human genetic and sequence-based physical maps', *Nature*, vol. 409, pp. 951-953.

Weller, J.I. 2001, 'Genetic mapping functions', in *Quantitative trait loci analysis in animals*, CAB International, New York, pp. 9-12.

©2009 The Australian Wool Education Trust licensee for educational activities University of New England

Glossary of terms

Annotated	Marked up - Identifying genes and other features in the DNA sequence
BAC	Bacterial artificial chromosome; an F plasmid engineered to act as a cloning vector than can carry large inserts[1]
Chromosome walking	A method for the dissection of large segments of DNA, in which a cloned segment of DNA, usually eukaryotic, is used to screen recombinant DNA clones from the same genome bank for other clones containing neighbouring sequences[1]
Clone contig	A set of ordered overlapping clones that constitute a chromosomal region or a genome[1]
cM	CentiMorgan: A unit of distance on a genetic or linkage map
Contig	A set of ordered overlapping clones that constitute a chromosomal region[1]
Cytogenetics	The cytological approach to genetics, mainly consisting of microscopic studies of chromosomes[1]
EST	Expressed sequence tag: a short sequence derived from mRNA sequence
Genetic map	A set of genetic markers that have been assigned to particular chromosomal locations based on linkage analysis / recombination frequencies
Interference	A measure of the independence of crossovers from each other[1]
Library	A collection of DNA clones obtained from one DNA donor[1]
Likelihood	Equal to the probability of observing a certain data set for given data values
Linkage map	Alternate name for a genetic map
Linkage phase	The arrangement of alleles of linked loci on chromosomes
Mapping function	A formula expressing the relation between distance in a linkage map and recombination frequency[1]
Mb	Megabase pairs: A unit of distance on a physical map equal to10^6 base pairs
Physical map	A map describing the physical structure of the genome. The ordered and oriented map of cloned DNA fragments on the genome[1]
Reverse transcription	Synthesis of a DNA strand from an RNA template[1]
RH map	Radiation hybrid maps: a physical map where distance between two marker loci is estimated from the number of breaks induced by radiation
Sequence assembly	The compilation of thousands or millions of independent DNA sequence reads into a set of contigs and scaffolds[1]
Sequence map	The DNA sequence of a chromosome or entire genome: the ultimate physical map
STS	Sequence tag site: a unique sequence of 100-500 bp in length, used as a marker in physical map construction
STS map	Sequence tag site map: a physical map derived from ordering clones into a contig based on overlapping sequence tag sites
Three-point testcross	A testcross in which one parent has three heterozygous gene pairs[1]

[1]Glossary terms taken from Griffiths et al. (2005).

©2009 The Australian Wool Education Trust licensee for educational activities University of New England