Practical 6. Genetic linkage and mapping.

Chapter · October 2018

CITATIONS
0

READS
2,244

3 authors, including:

Boris M. Sharga
Uzhhorod National University
60 PUBLICATIONS **85** CITATIONS
SEE PROFILE

Volodymir Feketa
Uzhhorod National University
58 PUBLICATIONS **4** CITATIONS
SEE PROFILE

Uzhhorod National University
Medical Faculty N2
Department of Fundamental Medical Disciplines

MEDICAL BIOLOGY PRACTICALS. GENETICS.
Practical 6. Genetic linkage and mapping.

By Boris M. Sharga, Diana B. Pylypiv, Volodymir P. Feketa

According to Mendel's 2nd law, the alleles of different genes, located on different chromosomes (chrs) have independent assortment (IA). The genes on the same chromosome (chr) may show no IA, if they are close together or ***linked***. They tend to stay together in meiosis (M) and usually passed to the same gamete (G). T.H. Morgan tracked this issue in *Drosophila melanogaster* [16]. He found, that *white eye* (*w*) and *miniature wing* (*m*) genes on *X* chr tend to be inherited together as linked. The linkage (L) was not complete. In F1 from the cross ♀♀*wm/wm* × ♂♂*w+m+/Y* all females were of wild type (*w+m+/wm*) and all males were with white eyes and miniature wings (*wm/Y*). When female and male flies of F1 were crossed, most frequent phenotypes in their progeny were grandparents' phenotypes: white eyes, miniature wings and wild type. These are coded by original genotypes of grandparents and called *parental* (Pr) *genotypes*, *parental classes (types)* or just *parentals*. 33% of the cross progeny comprised by flies with genes combinations that differ from Prs. They had white eyes and normal wings or red eyes and miniature wings. He called them *recombinant* (Rc) *classes*. To explain the data, T. H. Morgan suggested genes exchange during M between homology *XX* chrs in F1 females, as no exchange is observed in M between the non-homology *Y* and *X* chrs in males [18].

In similar experiment the pure line white-eyed, yellow bodied females (recessive traits, *wy*) were crossed with pure line wild (red-eyed, gray-bodied *w+y+*) males. Then, wild traits F1 females were mated with recessive traits F1 males. The progeny was obtained with only 1% of Rcs, the red-eyed, yellow bodied and white-eyed, gray-bodied flies. From these results T.Morgan concluded more close linkage for eye color and body color genes, than for eye color and wing size. The traits inherited together to a higher extent, the closer coding genes are on the chr [18].

The Rc/Rc and Prs/Prs ratios were close to 1 in these experiments, as the classes were reciprocal products of one exchange event. T.H.Morgan and E. Cattell introduced term '***crossing-over'*** for this exchange [17]. The *crossing-over* or *crossover* (CO) is a process of exchange of gene loci between *non-sister chromatids* of homologous chrs which may result in ***recombination*** (R), i.e., production of Rcs in *late prophase* of M I. CO breaks up the L between genes. In M II the genes that were linked will then assort independently and finally appear in different Gs.

The probability (P) of CO and thus, the P of genetic R increases with increase of *d* between 2 linked genes. The P of R serves as a measure of *d* between genes and allows the construction of ***genetic linkage map***, showing relative position of the genes and *d*s between them. This idea was first proposed by Alfred Henry Sturtevant, the undergraduate student, in 1911, when he and his teacher, T.H. Morgan, were discussing IA. During followed night A.H. Sturtevant built first L map [21].

The ***recombination fraction,*** *r* (or *q*) between genes *w* and *y* in the above study is: *r* = 0.01. The genes are ***linked*** as *r* < ½. If being *r* = ½, the traits are in IA (Mendel's 2nd law hold true). ***Recombination frequency***, RF (or θ) is the genetic L measure used in the L maps constrution. It is the frequency of a *single crossing over* (SCO) between 2 marker genes. The RF is a measure of recombination distance (*d*) between 2 genes. In example above 1% of RF (*r* = 0.01) was observed. It means that *d* between genes *w* and *y* is 1 map units (m.u.) or 1cM. A **centimorgan** (cM) is a unit describing a RF of 1% [5]. The map unit 1 **centimorgan** (cM) is used in honor of T.H. Morgan, 1 **morgan** = 100 m.u. = 100 cM.

The linked genes belong to the same L group. It is important to know which alleles on chrs are linked. We can depict the homological chrs genes in *coupled* (*cis*-) or in *repulsive* (*trans*-) positions (Fig. 1).

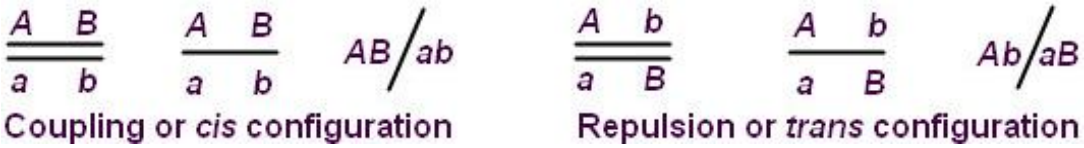

Fig. 1. The alternative positions of genes on chromosomes pair. *By D.B. Pylypiv.*

The Gs produced in M have alleles in Pr and Rc arrangement. If Pr configuration is *cis*-, so Rcs chrs must be in *trans*- position and *vice versa*. The RF between 2 mutant alleles is independent of whether they are on same chr (*cis*-) or if they are separated on different homologous chrs (*trans*-configuration). This was showed, particularly, in Morgan's study of *D. melanogaster X* chr linked *white eyes* (*w*) and *miniature wing* (*m*) genes (Fig. 2). In Cross 1 he allowed *white eyed, normal-wing* (*w* +) females and *red-eyed, miniature wing* (+ *m*) males to mate. Then, F1 *trans*-configuration females (*w* +/+ *m*) were crossed with males having both mutant alleles on *X* chr (*wm/Y*). The Prs and Rc classes were observed (Fig. 2, A).

The Cross 2 was done between *wm* gomozygous females and wild type males. Then, F1 females with 2 mutant alleles on the same chr were test-crossed. The Rc and Pr classes % were in quantities close to those in Cross 1. The difference between the classes was in the range of variation. The results of crosses 1 and 2 are not consistent with IA ratio 1:1:1:1. The 33.5 and 37.7% of *X* chr Rcs in Crosses 1 and 2 are RFs and these RFs < 50%. The Rc *X* chrs *wm* and ++ (Cross 1) or + *m* and *w* + (Cross 2) appeared in F1 females due to CO in M. These resulted in above calculated RFs for genes *w* and *m*. Thus, no matter, how mutant alleles are arranged: in *cis*- or *trans*- position the RF between them is the *same* [7].

The RFs are variable and depend from *d*s between linked genes allowing CO. So, the genes with RF < 50% are linked. No-linkage means 50% of COs. If genes are on different chrs or separated by large *d* on the same chr, they segregate as not linked, RF = 50%. A RF > 50% can't exist [3, 23]. Unlikely to most of living things, the males of *D. melanogaster* have no CO and R. So, genes of same chr location show complete L and they are inherited always as a group without R. For example, *cn* and *bw* genes for cinnabar or brown eyes are linked on chr 2. They are so distant, that in females show 50% RF. The cross of female *cnbw*/++ and male *cnbw*/*cnbw* results in offspring with Rc (*cn*+/*cnbw*; +*bw*/*cnbw*) and Pr (*cnbw*/*cnbw*; *cnbw*/++) traits in

1:1:1:1 ratio. In reciprocal cross produced Prs only, as there is no CO in male [7]:
P: ♀♀ ***cnbw/cnbw*** **×** ♂♂ ***cnbw/ ++***
G: ***cnbw,*** ***cnbw, +***
F1: ***cnbw/cnbw; cnbw/ ++*** **(1:1).**

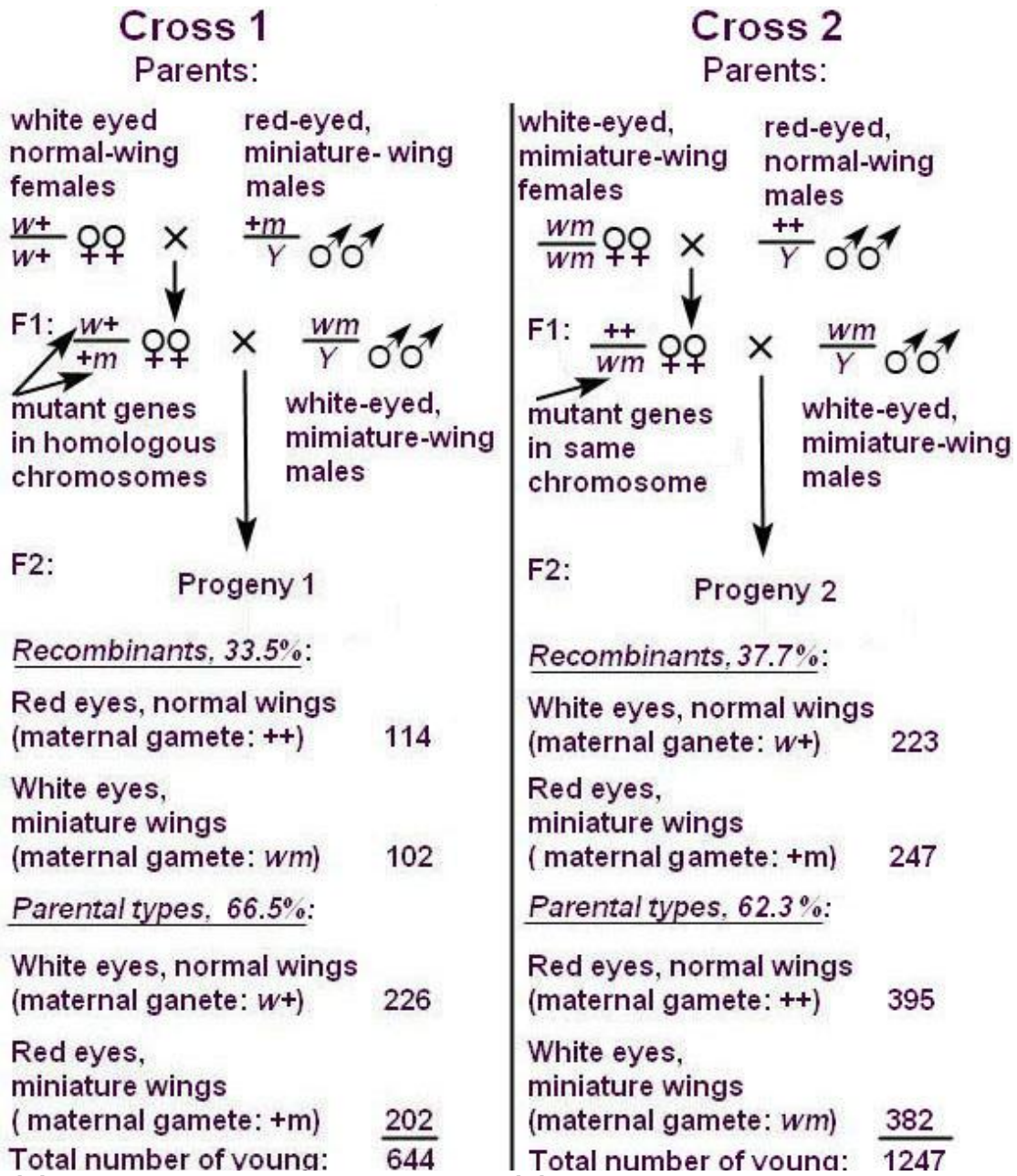

Fig. 2. Experiment, demonstrating that number of recombinants is independent from mutant genes configuration on chromosomes pair. *By D.B. Pylypiv.*

The COs are manifested by cross-like structures, the chiasmata, in prophase I of M. They were first found by F. Janssens at the Leuven University, Belgium in 1909 [1]. The chiasmata (in Fig. 3 as X) are produced from breakage and rejoining of chromatids, but only in non-sister chromatids they may result in R, as COs in sister chromatids exchange identical parts. When CO occurs, the parts of chromatids are exchanged, providing alleles R (Fig. 3, A). If one CO is followed by another CO

between the same chromatids, there is no R for marker genes, as second CO returns the markers to previous place (Fig. 4, A). If loci of the genes are very close (< 7 cM), a *double CO* (DCO) is very rare [13]. The higher *d*s increase the likelihood of a DCO. If the *d* between *A* and *B* is long enough, it allows more than two COs. When chiasma takes place outside the *A*-*B* region, e. g., between the gene *A* and centromere, the exchange by chr parts does not result in R, the genes remain in the same place (Fig. 4, B). During COs, the chromatids for exchange are joined by chiasmata randomly. No Rcs are produced in DCO involving 2 chromatids (2-strand DCO); 2 Rcs are produced when 2 exchanges have 1 chromatid in common (3-strand DCO); 4 Rcs are produced, when after 1st CO the 2nd CO involves the chromatids, which were not participating in 1st CO (4-strand DCO). The probabilities of 2-strand, 3-strand and 4-strand DCOs are ¼, ½ and ¼, respectively. The average of Rcs is: ¼ × 0 + ½ × 2 + ¼ × 4 = 2, i.e., on average 2 Rcs are produced from 2 chrs undergone DCO during M. As DCOs are undetected because of no R for particular marker genes, there is a difference in *d* measured in m.u. and by RF. The map *d* between two genes is a measure of CO. It equals half of average number of COs that take place between 2 marker genes during M. The RF between the genes tells how much of Rs were detected, it is a measure of R. The difference between map *d* and RF appears due to DCOs that do not yield Rs, so do not contribute to RF, but do contribute to map *d*. The difference is important only if *d* between 2 markers is large enough to provide possibility of DCOs. When *d* is *too short* to allow more than one CO, the RF and map units are the same, because there are no DCOs. On this basis 1 m.u. regarded as equal to 1% RF, because each CO at *d* < 7 cM results in R.

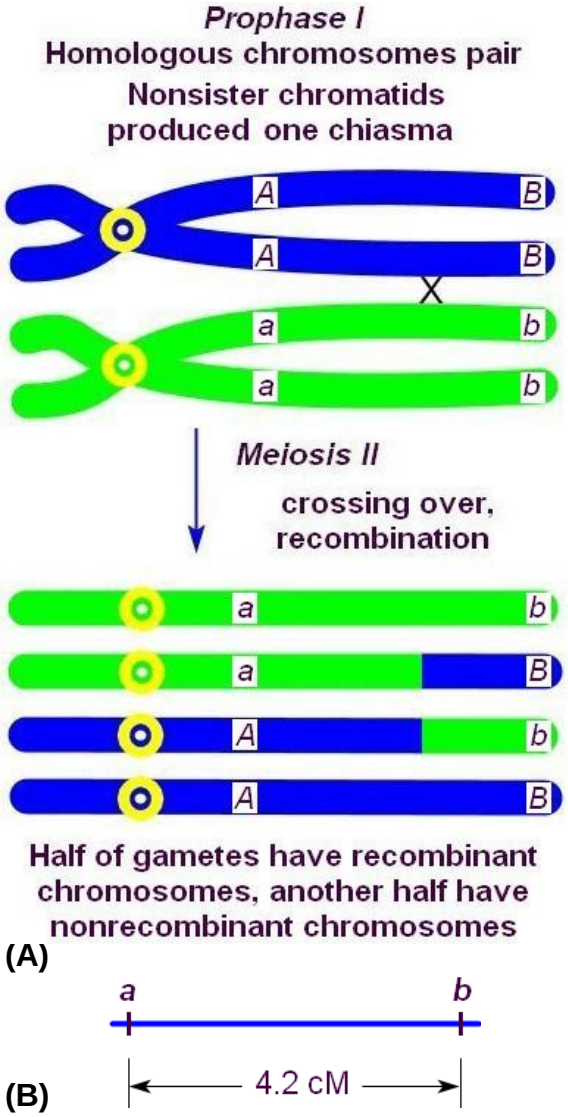

Fig. 3. Single crossover (A) and an example of genetic map (B). *By D.B. Pylypiv.*

The 1 m.u. is a length of chr in which on average 1 CO happened per every 50 cells undergoing M. If 1 of 50 meiotic cells has a CO, the frequency of CO is 1/50 × 100% = 2%. It is in correspondence with 1% of RF, because only half of chromatids in each cell with COs are actually *Rcs*. The *CO frequency* of 2% means, that 2 out of 200 chrs produced during M in 50 cells are Rc (Fig. 5). The 49 cells without COs produced 98*ab* and 98*AB* non-Rc chrs.

One cell with CO yields 2 Rc chrs (*aB* and *Ab*) and 2 chrs of Pr class (*AB* and *ab*). We can calculate the *r* as [7]:

$$r = \frac{1+1}{49+49+49+49} = \frac{2}{200} = 0.01 = 1\ map\ unit\ (m.u.) = 1\ centimorgan\ (cM).$$

The RF = 0.01 × 100% = 1%. Thus, 1% of RF means that 1 of 50 meiotic cells has a CO between the genes *A* and *B*. Independantly from *d* between two genes, the maximum RF between them is 50% and this is found for genes located on largest *d*s. The 50% RF is observed in genes situated on non-homologous chrs that have IA also.

The P of multiply CO increases with increase of *d* between genes and this complicates the mapping, because some of the multiply COs do not result in R and go undetected. Thus, true frequency of CO is higher, than detected by Rcs number.

The *d*s estimated on the basis of RF are *additive approximately*. For instance, if *d* between *a* and *b* is 4.2 m.u., the *b*-*c* region is 8 m.u., and the *A*-*C d* is 12.2 m.u., the gene *B* location must be between genes *A* and *C*.

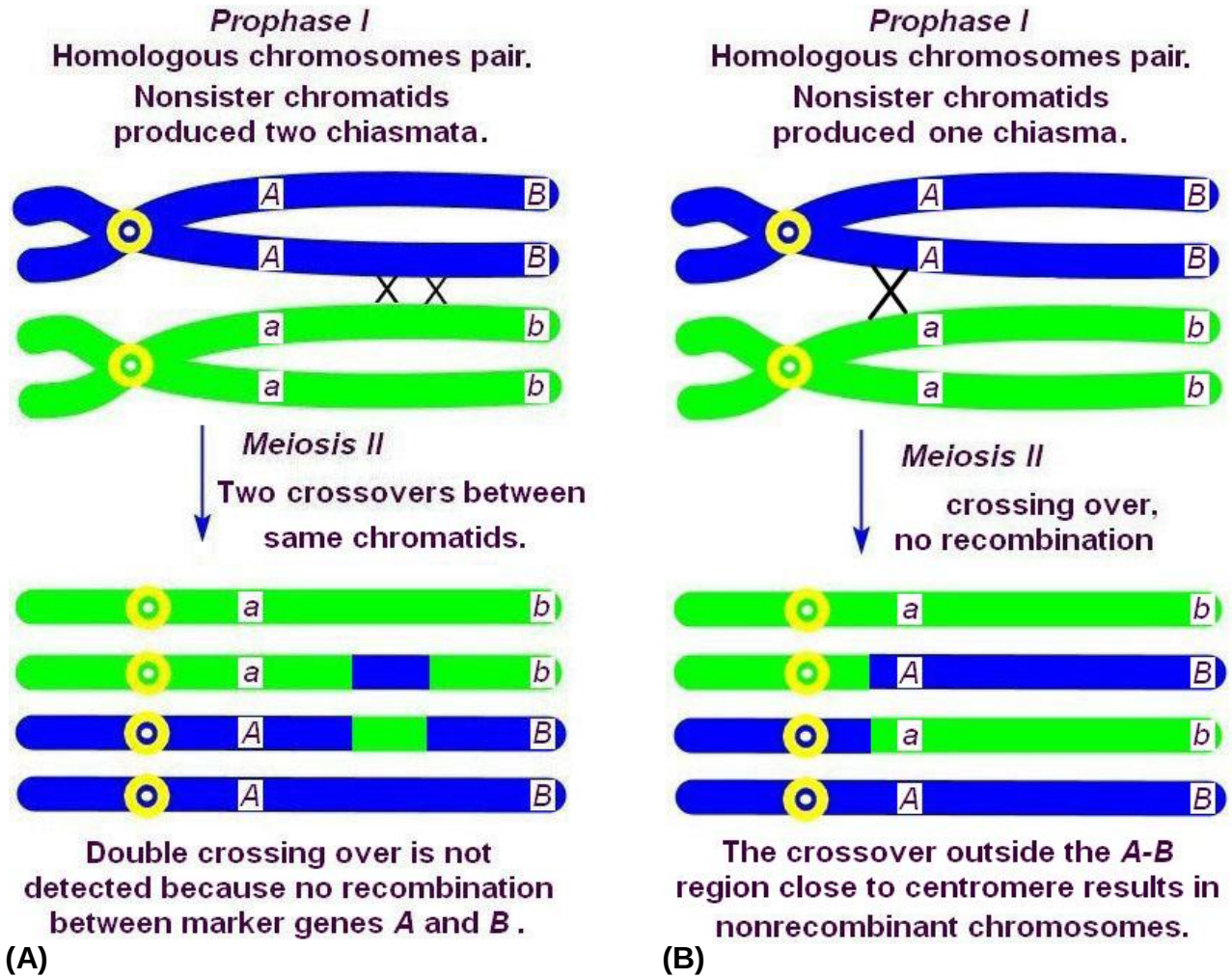

Fig. 4. The crossovers resulting in no recombination. *By D.B. Pylypiv.*

To avoid the multiply COs effect, we must build the map on RF data for closely linked (≤ 10 m.u. apart) genes. If the Rc and Pr classes are 50%, the *d* = 50 cM. The correlation between physical distance and *d* is not strict. Physically chrs and orders of the genes on them are same, but the map *d*s in chrs of two sexes may differ. The minimum RF is 0%. The *d* = 0 between any of two genes on chr of fruit fly males,

as there is no CO and R in this sex. These genes are in IA for the fly males.

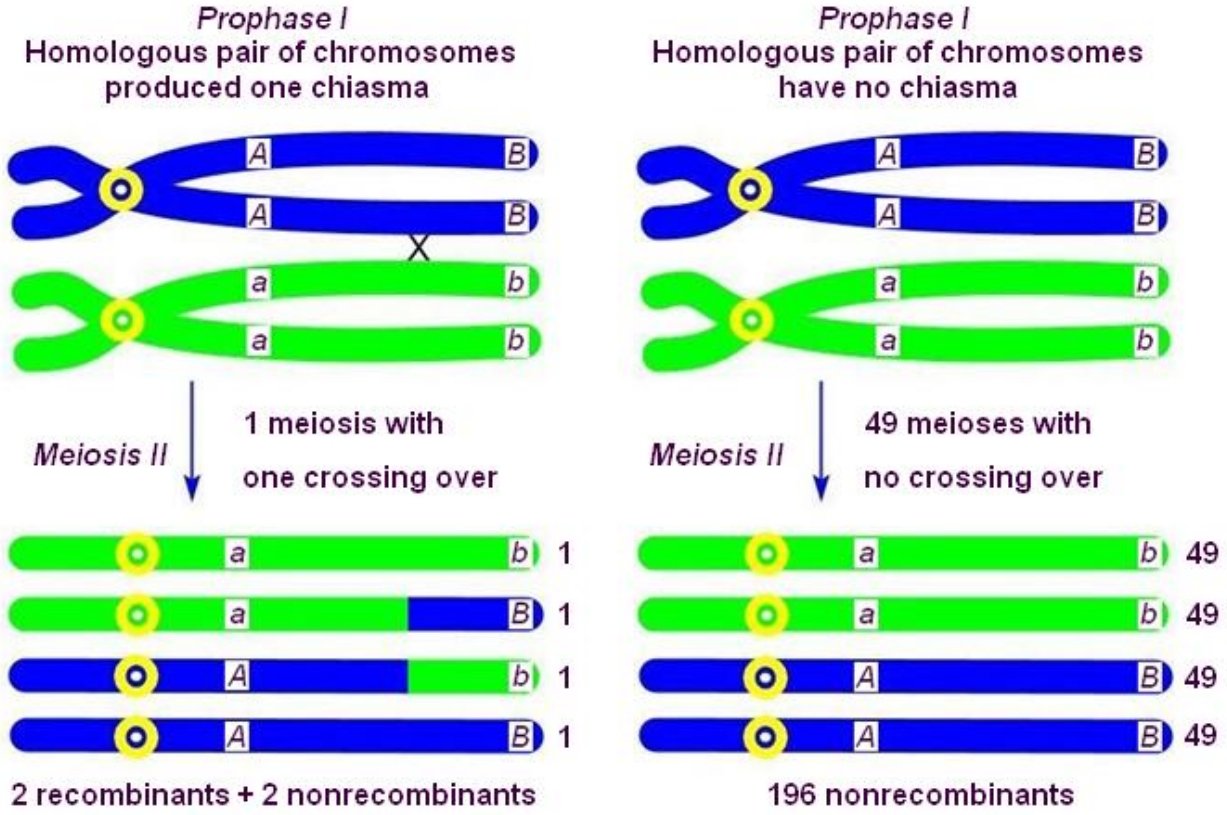

Fig. 5. Chromosomes outcome in 50 meiotic cells, in which 1 has crossover. *By D.B. Pylypiv.*

Meiotic R differs in two sexes of most mammals [6, 8, 15, 19], e.g., humans. The genome-wide R levels are higher in females than in males, and the placement of exchanges varies between the sexes. The R differences in human sexes may derive from chromatin differences established prior to the onset of the R pathway [6].

The heterochromatin on both sides near centromere is poor for genes. It has low frequencies of CO and R. The large physical distance here may correspond to 10 times smaller *d* in L map. The euchromatin has high CO and RFs, these parts of chr physical distances correlate better with *d*s on L map. Note, that even COs number decrease the *d* [7].

The L map on Fig. 6, shows only genes of interest. It is a line with 3 marks for *a*, *b* and *c* genes separated by particular *d*s, e.g., 4.2 cM for *a*-*b*; 8 cM for *b*-*c* and 12.2 cM for *a*-*c*. The maps with *a* on the left, *b* in the middle and *c* on the right or in opposite orientation are right for these data. Having RF for an additional gene *d* we can find the true *a* and *c* sites. If RF between genes are 3,8 for a-d, 8 for b-d, 16 cM for c-d, the true genes order *dabc*. RF-based L map is not precise at large *d*s.

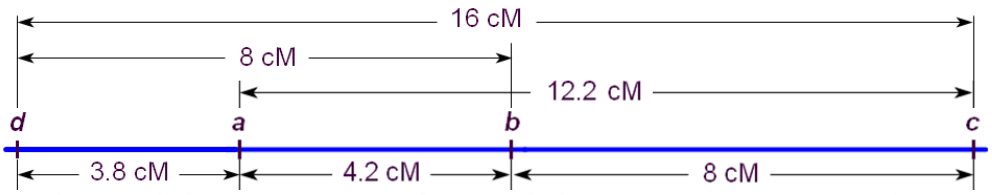

Fig. 6. Linkage map of *a*, *b*, *c* and *d* genes. *By D.B. Pylypiv.*

Two-point test-cross, the test-cross between 2 genes, allowing construction of L map on the basis of their RF. Presume, that two-point crosses for genes *a*, *b*, *c*, *d* yield the RFs showed in Table 1. The RF for each pair of genes *a*-*b*, *a*-*c*, *a*-*d* is 50%.

Table 1. Two-point crosses results

Gene loci in test-cross	R frequency (%)
a and *b*, *a* and *c*, *a* and *d*	50
b and *c*	15
b and *d*	13
c and *d*	26

These pairs may therefore be on different chrs or on large distance on the same chr. We must place them in separate L groups. The genes *b* and *c* or *b* and *d* are linked through region of 15 and 13 m.u. (Fig. 7), as they have RFs 15 and 13%, respectively. The *a* and *d* genes belong to different L groups, as their RF is 50%. To estimate, whether the *d* is 13 m.u. to the right or left of gene *b*, we need regard the *c*-*d* distance. If gene *d* is 13 m.u. to the right the *b*-*c* map fragment must be 2 m.u. This is not in compliance with the cross RF data. Thus, the gene *d* is 13 m.u. to the left of gene *b*, and genes *d* and *c* is 15 + 13 = 28 m.u. apart. The RF obtained for these 2 genes from the test cross is 26 m.u. The difference is due to DCOs without R between these genes. As they are not counted, the RF is underestimated. The rule is true: *RF is approximately additive*. The map from two-point cross is not precise, as some DCOs are missed.

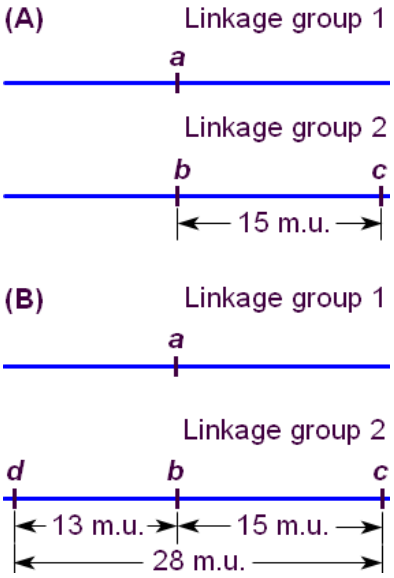

Fig. 7. Map construction.
By D.B. Pylypiv.

Three-point test-cross is a test-cross between 3 genes for mapping. The positions of the 3 genes are estimated in offspring, when SCOs and some DCOs are indicated, allowing more precise map *d*s estimation.

The cross *AABBCC* × *aabbcc* was performed. Next, the F_1 *AaBbCc* female was test-crossed to an *aabbcc* male. Three CO variants are possible for *ABC* genes with Rc chrs production: SCOs (between *A* - *B* or *B* - *C*) and a DCO (between *A* - *B* and *B* - *C*), Fig. 8, A. The G composition of F1 female reflects the progeny classes (Table 2), as male in test-cross is complete recessive homozygote. How estimate the *d*s and genes order? First, determine the Prs, they are most frequent in progeny. Here the *ABC* and *abc* are Prs, as they are most numerous. Second, find DCOs. They are always in the lowest number. Here *ABc* and *abC* types are in the lowest number. Remember, that in DCO progeny only middle gene differs from non-Rcs. From Pr and DCO classes comparison (*ABc* and *ABC* or *abC* and *abc*), the *c* is "switched" as middle gene. The *acb* is true genes order and the Fig. 8, A can be rearranged as Fig. 8, B.

Third, determine the L *d*s *A*-*C* and *C*-*B* by dividing the total number of Rc Gs onto the total number of Gs including DCOs. The DCOs frequencies are used in the calculations of both *d*s. So, the *d* between *A* and *C* is (68+70+3+2)/1000 × 100 = 14.3 cM and the *d* between *C* and *B* is (27+25+3+2)/1000 × 100 = 5.7 cM. Now, the L map can be constructed, that shows the *d*s between the genes (Fig. 8, C). But, their precise position on chr and which chr they are on remains unknown.

Using cytogenetic analyses [20], the scientists correlate the L maps with

particular chrs and with specific regions on them. We will learn these methods later.

Usually, first CO affects the P of the next CO in adjusted region. This interaction is called ***interference*** *(I)*. If the COs in *A-C* and *C-B* regions are independent, the P of double Rs is a product of RFs in these regions. The *r* for *d A-C* is 0,143, and *r* value for *C-B* region is 0.057, so if DCO are independent, they are expected at the frequency: 0.143 × 0.057 = 0.0082. In the offspring of 1000 individuals 8 double Rs are expected: 0.0082 × 1000 = 8.2≈ 8. In progeny we have only 5 DCO Rs suggesting that two regions are not independent, they are interfering.

Table 2. Three-point test-cross results

Genotype	Observed	Total	True order	Gametes Type
ABC	403	1000	*ACB*	Pr
abc	402		*acb*	Pr
AbC	25		*ACb*	S- crossover between *C* and *B*
aBc	27		*acB*	S- crossover between *C* and *B*
ABc	2		*AcB*	Double- crossover
abC	3		*aCb*	Double- crossover
Abc	70		*Acb*	S- crossover between *A* and *C*
aBC	68		*aCB*	S- crossover between *A* and *C*

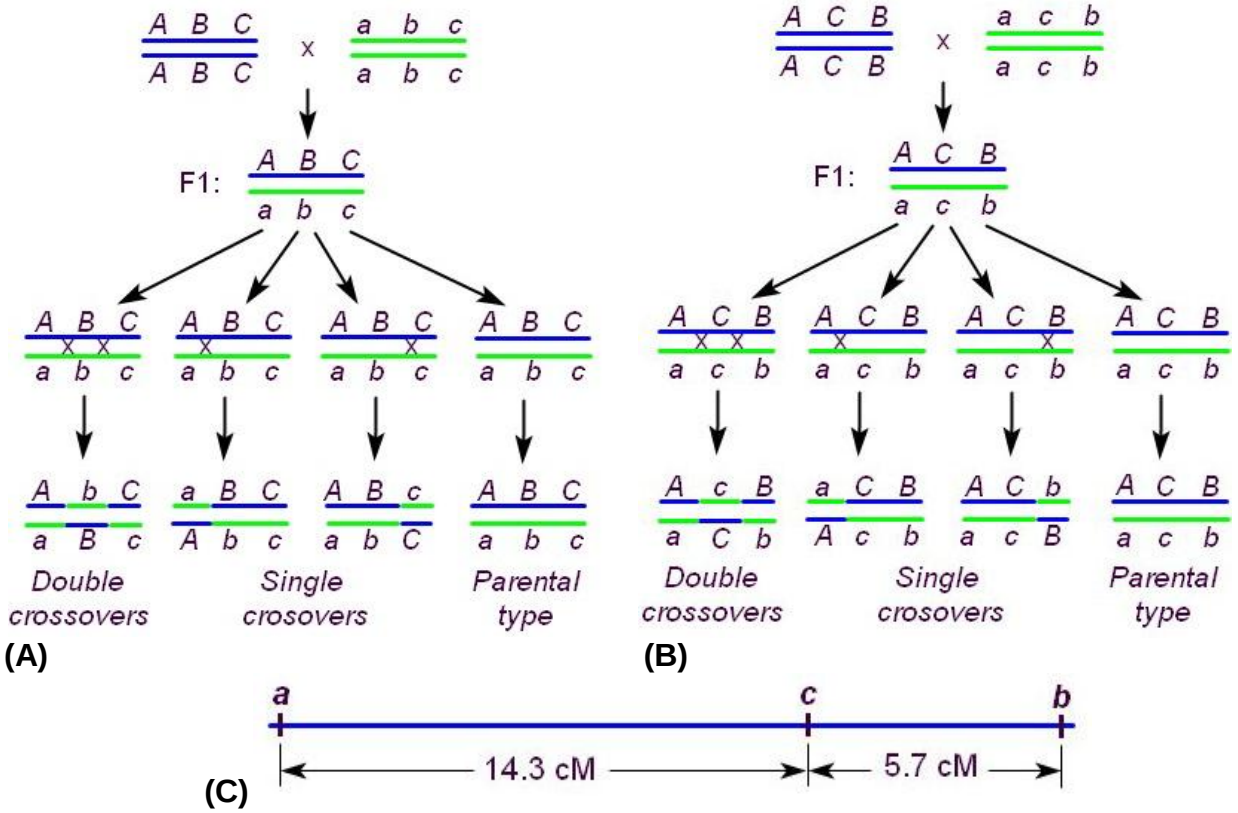

Fig. 8. Gametes at expected (A) and true order of genes (B) and L map derived in three point cross after true order of genes was estimated (C). *By D.B. Pylypiv.*

The **interference** *(I)* is evaluated by subtracting of the **coefficient of coincidence (c. o. c.)** from 1: *Interference (I) = 1- c.o.c*. The c.o.c. is a ratio of observed to expected DCO Rcs or ratio of their frequencies. For this experiment:

$$Interference\ (I) = 1 - \frac{Observed\ DCO\ number\ or\ RF}{Expected\ DCO\ number\ or\ RF} = 1 - 5/8 = 0.375 \text{ or } 37.5\%.$$

For regions, where never DCOs are observed, the c. o. c. = 0, thus $I = 1$ or 100%. In most cases the I values are between 0 and 1. However, in some studies observed DCO Rcs are more numerous than expected, and $I < 0$.

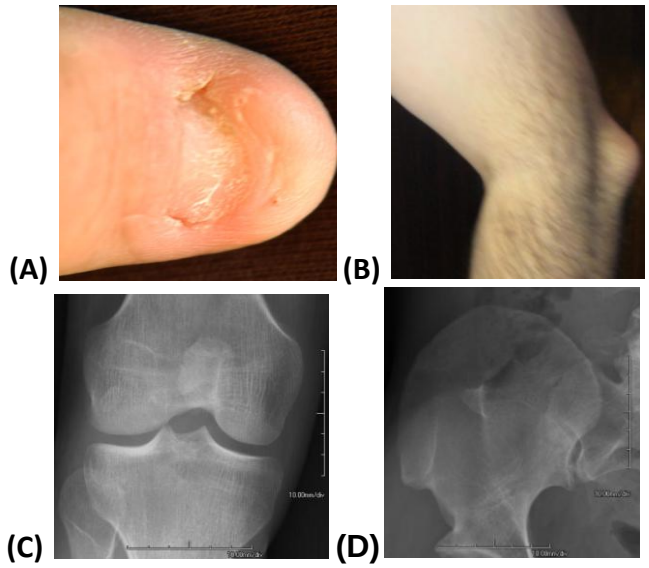

Fig. 9. Symptoms of nail-patella syndrome: Thumbnail (A), elbow (B) and AP radiographs of hypoplastic patella (C) and iliac crest with a bony exostosis or posterior iliac horn (D).
(A) By J.Wijten [9], (B) By J.Line [10],(C) and (D) by Jto410 [11,12].

Human genes mapping is complicated by small number of progeny in families and by the inability to have needed mating. Thus, data from several pedigrees are usually combined for analysis. The L between the nail–patella syndrome (NPS) and *ABO* blood types loci was among the first studied in human. NPS is rare dominant disease (2/100,000 births). The key features of NPS are poorly developed nails, patellae (kneecaps) and the presence of iliac horns (Fig. 9). NPS results from the loss of function or mutations in the LMX1B gene. These may results in dorsalising signals reduction, causing poor dorsal structures (e.g., nails and patellae) development [24]. Other abnormalities are: elbow deformities, abnormal shape of pelvic (hip) bones and kidney disease.

The ABO blood groups and NPS loci L was studied in pedigree whith both traits segregation. NPS is manifested, when at least 1 mutant allele *N* is present (Fig.10). The parent I1 is heterozygote (*Nn*). All diseased persons in generation II have either blood type B or O. In generation III sons 2 and 5 have a blood group A and NPS genotype and persons III1, III3 and III4 are healthy Rcs. So, there are 3 Rcs among 7 children in row III. This gives us $r = 3/7 = 0.4286$ and a d of 42.86 cM. We might conclude, that the loci for NPS and blood types are linked with the RF = 42.86% < 50% for IA. However, we must regard the possibility that the genes are in IA also. To verify their L, we can calculate the $\log_{10}$ of odds (***lod***) ***score.*** First, it is necessary to estimate the ratio of the P to have the observed results with L to the likelyhood to have them with IA. Second, calculate the log_{10} of this ratio, the *lod score* [13]. The P of 2 Rcs types: 0.4286/2 = 0.2143. The P of 2 Pr types: (1- 0.4286)/2 = 0.2857. The P of a given birth sequence (4 Prs + 3 Rcs) is the product of each of the independent events Ps: $(0.2857)^4(0.\ 2143)^3 = 6.6625 \times 10^{-3} \times 9.8416 \times 10^{-3} = 6.557 \times 10^{-5}$. If two loci are unlinked, the $r = 0.5$ and P of any genotype would be 0.25. The P of the birth sequence based on no linkage would be $(0.25)^7 = 6.1035 \times 10^{-5}$. Now, the ratio of the L likelyhood and the non-L P is: $6.557 \times 10^{-5} / 6.1035 \times 10^{-5} = 1.0743$ and the *lod* score is: $log_{10}1.0743 = 0.0311$. How to interpret the *lod* score? Positive *lod* score supports a linkage. A *lod* score ≥ 3 is an evidence for L.

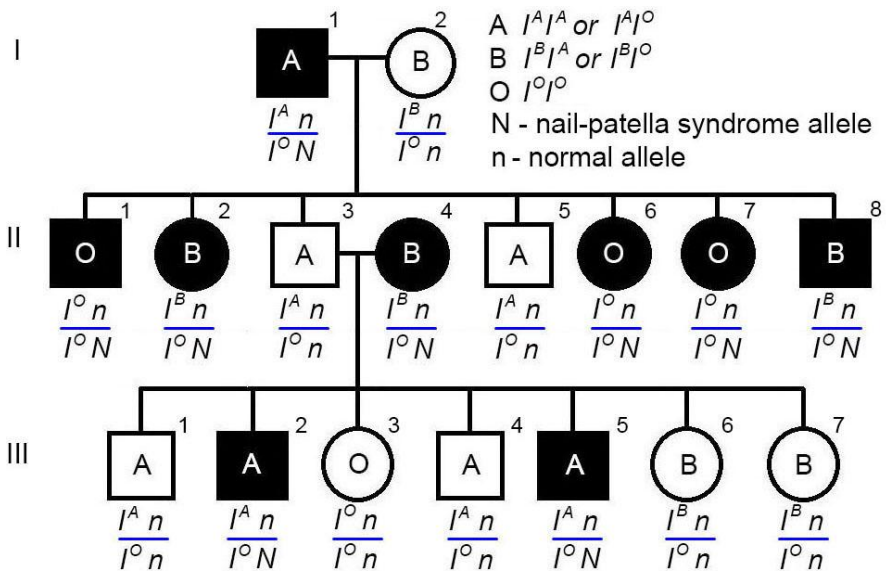

Fig. 10. ABO blood types and nail–patella syndrome L
ABO blood type and genotype are indicated inside and below each circle or square, respectively. The individuals 1, 3, and 4 in row III are recombinants. *By D.B. Pylypiv.*

The not informative is *lod*=0. The *lod* value ≤ -2.0 is non-L evidence. The *lod* score >-2< 3, like in regarded case (Fig. 10), is inconclusive and more data is needed from additional families. On the *lod* score 0.0311 we can't decide, if NPS and *ABO* blood groups loci are linked or not. Usually, several pedigrees are used to get the estimates for *r*. *Lod* scores are additive across several pedigrees. The largest *lod* score is correspondent to one of *r* estimates and to most likely *d*.

Presume, the P of our results in L is 0.1 and the P of these with IA is 0.0001. So, the ratio 0.1/0.0001 = 1000 and the *lod* score = $log_{10}1000 = 3$. This tells, that is 10^3 times less likely to obtain these data in IA than from L.

The *d*s are not additive in many cases. To avoid redoing the L map each time when new loci are found, the *d*s are mapped using a *mapping function* (MF). The MF are *I*-dependent. With *complete I* or small *d*s, a MF is: $d = r$. With *no I*, the *Haldane* MF is used, $d = - ½ ln(1-2r)$, and given the map *d*, the $r = ½ (1-e^{-2d})$. *Kosambi's* MF allows some *I*: $d = ¼ ln[(1+2r)/(1-2r)]$ [22]. MFs of Haldane, Kosambi, Carter-Falconer or Morgan help to estimate the *d* in cM from *r* data more accurately [1, 2, 3, 4, 14, 22]. Precise *d*s are important for the construction of large-scale L maps. Softwares, such as QTL Cartographer, Linkage, Mapmaker, etc., enabling us to calculate different MFs and provide comparative analysis while producing L maps.

The genes-markers are coding for easily detectable phenotypic traits. The shortage of these of traits in humans is among main obstacles for mapping.

Molecular markers mapping utilizes molecular biology methods, particularly, restriction fragment length polymorphisms (RFLPs), variable number of tandem repeats (VNTRs), single nucleotide polymorphisms (SNPs) and modern sequencing to provide wide variety of molecular markers and use them in L analysis similarly to use of traditional markers. About these and other molecular tools we will learn later.

Solved problems

Problem 1. Two homozygotes are crossed and the F1 heterozygote is crossed back with the double recessive individual. The progeny was as follows: 510 *a+b+*, 490 *ab*, 45 *a+b* and 55*ab+*. Estimate L between *a* and *b*. Draw the *a-b* L map.

Solution: The *a+b+* is wild, *ab* is mutant classes. Parents: *a+b+/a+b+* × *ab/ab*. F1 back cross: *a+b+/ab*×*ab/ab*. The progeny of back cross: 510 *a+b+*,

490 *ab*, 45 *a+b*, 55 *ab+*. Prs: *a+b+* 510 + *ab* 490 = 1000. COs: *a+b* 45 + *ab+* 55 = 100. % of CO 100 /1000 × 100% = 10 %. L map for genes *a* and *b*: *a* 10% *b*.

Problem 2. White endosperm in corn (*p*) is recessive to purple (*P*) and shrunken (*f*) is recessive to full (*F*) kernel. A pure white full plant was crossed to a pure purple shrunken. The F1 is then test crossed and the progeny was as follows: white full –557, purple shrunken – 443, white shrunken – 55, purple full – 45. Calculate the *d* between *p* and *f*. Draw a L map.

Solution: It is a two point test cross. Parents of the cross = *Pf* ×*pF*. Pr – purple shrunken = 443, white full – 557. Rcs – white shrunken – 55, purple full – 45. Total offspring – 1000. % of CO between *p* and *f* = 55 + 45 ×100 /1000 = 10 %. L map = *p* 10 % *f*.

Problem 3. Forked bristles (*f*), miniature wings (*m*) and white eyes (*w*) are sex linked in *D. melanogaster* and recessive to the wild traits, the straight bristles (*f+*), long wings (*m+*), red eyes (*w+*). F1 females from a cross *wfm/wfm* × *w+f+m+* were mated with *wfm* males and gave the young, %: red, straight, miniature - 2.5; white, forked, long - 2.5; white, straight, miniature- 7.2; red, forked, long - 7.2; white straight long - 15,5; red, forked, miniature-15,5; white, forked, miniature – 25; red, straight, long – 25. Estimate Pr, SCO and DCO types. Verify the gene order. Determine the % of CO between the genes, the c.o.c. and *I*. Construct the L map.

Solution: Pr types are, %: white, forked, miniature *wfm* - 25 and red, straight, long, *w+f+m+* – 25. The least numerous types are DCOs, %: white, forked, long, *wfm+*- 2.5 and red, straight, miniature, *w+f+m* - 2.5. From comparison of Pr and DCO types, the *wfm* and *wfm+* or *w+f+m+* and *w+f+m* we can see that *m* gene is 'switched' as a middle gene. So, the true position of genes is *wmf*.

White long straight, *wm+f+* - 15,5%. Red, miniature, forked, *w+mf* - 15,5%.
% of SCO between white and miniature: 15.5 +15.5+ 2.5+ 2.5= 36. White, miniature, straight, *wmf+* - 7.2%. Red, long, forked, *w+m+f* – 7.2 %. The % of SCO between miniature and forked: 7.2 + 7.2 + 2.5 + 2.5 = 19.4. Observed DCO = 5%. Expected DCO = 36/100 × 19.4/100 × 100 = 0.36 × 0.194 × 100 = 6.984%. C.o.c. = observed DCO /expected DCO = 5/6.984 × 100 = 71,59 ≈ 72%. Interference, *I* = 1- 0.72 = 0.28 = 28%, i.e., we noted 72% of DCOs that were expected and 28% of expected DCOs did not form due to *I*. L map is: *w* 36% *m* 19.4% *f*.

Problem 4. The genes order d–e–f has *lod*=0 and order Ad–Ae–Af has a negative *lod* score of –2. What does it mean?

Solution: That means that it is 10^2 times more likely to obtain the observed data from the first order of genes than from their second order.

Problem 5. A testcross *AaBb*×*aabb* produced in total 200 offspring of the following types: *AaBb* - 64, *aabb* - 76, *Aabb* - 29 and *aaBb* - 31. Provide conclusion concerning the L or IA of the genes.

Solution: To test for IA or L between two genes, we need to state the *null hypothesis* "Luck of L"which yields the ratio 1:1:1:1. Then, calculate a series of three *chi-square tests* (X^2*test*) and estimate the probability *p*, that allows accept or reject the *null hypothesis* (H_o). From separate examination of each locus we can see that the observed numbers (*O*) differ from the expected numbers (*E*). For locus *A* the cross *Aa* × *aa* was expected to yeld ½ *Aa* and ½ *aa*. Instead for these we observed 64 + 29 = 93 of *Aa* and 76+31=107of *aa*. For locus *B* the cross *Bb* × *bb* was expected to yield

½ *Bb* and ½ *bb*. The *Bb* were produced less than expected 64+31= 95 and *bb* were observed in more large quantity, than expected 76+29 = 105.

Applying the $X^2 test$ to these *O* and *E* data, we calculate for locus *A*:

$$X^2 = \sum \frac{(O-E)^2}{E} = \frac{(93-100)^2}{100} + \frac{(107-100)^2}{100} = \frac{49+49}{100} = 0.98,$$

and for the locus *B* :

$$X^2 = \sum \frac{(O-E)^2}{E} = \frac{(95-100)^2}{100} + \frac{(105-100)^2}{100} = \frac{25+25}{100} = 0.5$$

The digree of freedom (df) for loci *A* the *B* is n-1= 2-1 =1, where *n*, the number of classes equals 2, because 2 classes of progeny are expected for either of loci.

Looking up chi-square values 0.98 and 0.5 in Table 3, we found that the calculated probability *p* associated with these X^2 values is above 0.05 (a critical probability for rejecting the null hypothesis) and calculated X^2 (0.98 and 0.5) are below the critical X^2 value 3.841 at the degree of freedom 1. So we can conclude that there is no significant difference between the 1:1 ratio that we expect in the offspring of the testcross and the ratio that we observed.

Now we can test the independent assortment of genes at the two loci. The total number of progeny is 200 and 4 classes are produced, thus 50 individuals expected in each. The *O* and *E* numbers of progeny can now be compared in chi-square test:

$$X^2 = \sum \frac{(O-E)^2}{E} = \frac{(64-50)^2}{50} + \frac{(76-50)^2}{50} + \frac{(29-50)^2}{50} + \frac{(31-50)^2}{50} = 33.48$$

Here, we have 4 classes of phenotypes; so the df = n-1 = 4-1 = 3. Calculated X^2 value 33.48 is above the X^2 value 7.815 at critical *p* = 0.05 and df = 3 (see Table 3). The *p* corresponding to calculated X^2 is much lower than 0.05. So, reject the null hypothesis and conclude that genes are not in IA and must be in L.

Table 3. Critical values of the X^2 distribution

df \ *p*	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005
1	.000	.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278

Problem 4. You are medical geneticist in research on 3 rare recessive human mutations (*s*, *c* and *g*) which are also present in mice. One student has produced a genetic map of these genes (*s* 6.8% *c* 4.2% *g*) based on his cross of wild pure line animals with homozygotic recessive for these traits. Thinking that student was not accurate in his work, you have repeated his experiment. The progeny of your cross is *SCG* - 656, *scg* - 620, *Scg* - 83, *sCG* - 85, *scG* - 72, *SCg* - 68, *sCg* - 7, *ScG* - 9 (in total 1600). Is his map the same as yours?

Solution: Compare data of the studies and calculate the expected data based on the student's map. Your data are used as observed values in a X^2 test. Of 1600 progeny, 6.8% are Rcs between *S* and *C*: 1600 × 0.068 = 108.8 SCOs + DCO Rcs. The 4.2% are Rbs between *C* and *G*, thus 1600 × 0.042 = 67.2 SCOs + DCO Rcs. The DCO Rcs are: 1600 × 0.068 × 0.042= 2.5792 ≈ 4.6. Thus, *S*-*C* SCO Rcs are 108.8 – 4.6 = 104.2 *C*-*G*; SCO Rcs are 67.2 – 4.6 = 62.6 and Non-Rcs are: 1600 - 104.2 - 62.6 - 4.6 = 1428.6. The calculated classes are 4 pairs of expected reciprocals. So, divide these numbers by 2. Now, use X^2 test to see wether the calculated numbers are consistent with your data (Table 4). The null hypothesis (H_o) is: 'no significant difference between observed and expected values'. From your data, the SCOs between *S* and *C* loci are: *Scg* - 83 and *sCG* - 85. The SCOs between *C* and *G* loci are: *scG* - 72 and *SCg* – 68. The DCOs from the cross are: *sCg* – 7 and *ScG* – 9 animals. % of CO between *S* and *C*: (83+85+7+9)/1600×100% = 11.5%. % of CO between *C* and *G*: (72+68+7+9)/1600×100% = 9.75%. The L map for your data is: *s* 11.5% *c* 9.75% *g.* It is quite different from the student's map.

Table 4. Chi-square test result

Phenotype	O	Total	E	O-E	$(O-E)^2$	$(O-E)^2/E$	
SCG	656	1600	714.3	- 58.3	3398.89	4.76	$X^2 = \sum(O-E)^2/E =$ 181.38. Table 3: the critical X^2 =14.067 at df 7 and p = 0.05. X^2=181.38> 14.067 and $p < 0.05$. H_o is rejected.
scg	620		714.3	- 94.3	8892.49	12.45	
Scg	83		52.1	30.9	954.81	18.33	
sCG	85		52.1	32.9	1082.41	20.77	
scG	72		31.3	40.7	1656.49	52.92	
SCg	68		31.3	36.7	1346.89	43.03	
sCg	7		2.3	4.7	22.09	9.6	
ScG	9		2.3	6.7	44.89	19.52	

Problems for individual work of students

Problem 1. Homozygous *Drosophila* line has the autosomal recessive genes *a*, *b*, *c* linked in that order. The *abc/abc* females were crossed with *a+b+c+/Y* males. Then the F1 males were crossed with their *abc/abc* sisters with production of the F2 progeny: *a+b+c+* 555, *abc* 445, *abc+* 57, *a+b+c* 43, *ab+c+* 47, *a+bc* 44, *ab+c* 5, *a+bc+* 4. Give the RF between *a, b* and *c*, the c. o. c., *I* and a L map for *a, b* and *c*.

Problem 2. A pure line wild fruit fly was crossed with a homozygous (*cu*, up curved wings), claret (*ca*, claret eye colour), fluted (*fl*, creased wings) fruit fly. F1 females were test-crossed with production of offspring: fluted – 5, claret curled – 5, curled – 45, fluted claret – 55, claret – 182, fluted curled -178, fluted claret curled – 395, wild type – 405. Give the true genes order, c.o.c., *I*, map *d*s between the genes.

Problem 3. In *B. mori* white-banded wing, *wb* and red eyes, *re* genes are recessive to wild re^+ and wb^+ genes, which are on the same chr. The insects *wb re* and *wb+re+* are crossed. All F1 have normal eyes and wings. The F1 are test-crossed with *wb re* moths. The progeny is: 459 *wb+re+*, 14 *wb+ re,* 16 *wb re+,* 441*wb re*. What is the map *d* between *wb* and *re*? What proportions classes would be expected if the genes *wb* and *re* were located on different chrs?

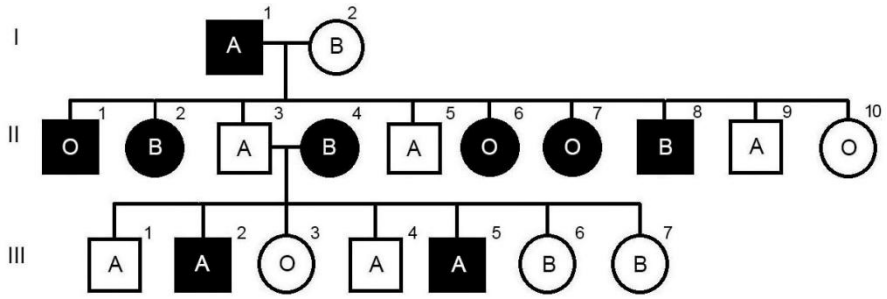

Fig. 11. Nail-patella syndrome and blood groups linkage
By D.B. Pylypiv.

Problem 4. What are the genotypes in the pedigree with NPS, Fig. 11? Calculate the *lod* scores and estimate the linkage relationship between NPS and *ABO* blood group loci.

Literature

1. Carter T.C. Falconer D.S. Stocks for detecting linkage in the mouse and the theory of their design// J. Genet.- 1951.- Vol. 50.- P. 307–323.
2. Casares P.A. Corrected Haldane's map function to calculate genetic distances from recombination data// Genet.- 2007.- Vol. 129. - N3.- P. 333-338.
3. Castle, W. E. Is the arrangement of the genes in the chromosome linear?// Proc. Nat. Acad. Sci. USA, 1919.- Vol. 5, N2.- P.25–32.
4. Chhatre V.E. How to calculate recombination fraction from Kosambi distance and *vice versa*. //http://www.crypticlineage.net/lib/Kosambi.pdf
5. Griffiths A.J.F., et al. Introduction to Genetic Analysis.- 11th Ed.- N.Y., W.H. Freeman & Company, 2015.- 868 p.
6. Gruhn J.R. et al. Cytological studies of human meiosis: sex-specific differences in recombination originate at, or prior to, establishment of double-strand breaks//Plos One.-2013.-Vol. 8,N12–P.1.
7. Hartl D.L. Essential genetics. A genomics perspective. 6th ed.- Jones & Bartlett Learning, LLC.- New York.- 2014.- 572 p.
8. Hassold T. et al. Cytological studies of meiotic recombination in human males//Cytogen. Gen. Res. - 2004.- Vol. 107, N3-4.- P. 249–55.
9. http://www.janwijten.com
10. https://commons.wikimedia.org/wiki/File:Nail-patella_syndrome_(NPS)_Elbow1.JPG
11. https://en.wikipedia.org/wiki/File:Patella_radiograph_in_Nail_Patella_Syndrome.jpg
12. https://en.wikipedia.org/wiki/File:Posterior_iliac_horn.jpg
13. Jugal Gogoi, Tazid Ali Estimation of genetic recombination frequency with the help of logarithm of odds (LOD) method// IJIRSET.- 2015.- Vol. 4, N 8.- P. 6768-6772.
14. Kosambi D.D. The estimation of the map distance from recombination values//Ann. Eugen.- 1944.- Vol. 12.- P.172–175.
15. Lenormand T. The evolution of sex dimorphism in recombination//Genet.-2003.-Vol.163.-P. 811.
16. Morgan T. H. A modification of the sex ratio, and of other ratios, in *Drosophila* through linkage// Z. Indukt. Abstam-mungs. Vererbungsl.- 1912.- N7.- P.323-345.
17. Morgan T.H., Cattell E. Data for the study of sex-limited inheritance in drosophila// *J. Exp. Zool*.- 1912.- Vol. 13.- P.79.
18. Morgan T.H., et al.The mechanism of mendelian heredity.- Henry Holt & Co, N.Y., 1915.-262 p.
19. Petko P.M., et al. Crossover interference underlies sex differences in recombination rates //Trends Genet.- 2007.- Vol.23.- P.539.
20. Ruddle F.H., Kucherlapati R.S. Hybrid cells and human genes//Sci. Amer.-1974.-Vol.231,N1.- P.36 –44.
21. Sturtevant A.H. The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association//J. Exp. Zool.- 1913. – Vol.14.- P.43–59.
22. Tan Y. D. , Fornage M. Mapping functions// Genet.- 2008.- Vol.133.- P. 235-246.
23. van der Werf J. Ch 5. Basics of linkage and gene mapping.– P.45-54. https://jvanderw.une.edu.
24. Wright M.J. et al. Achondroplasia and nail-patella syndrome: the compound phenotype// J. Med. Genet.- 2000.- Vol. 37, N9.- P.25.