Animal Genetics & Breeding

PRINCIPLES OF ANIMAL AND POPULATION GENETICS

(UNIT - II)

- Alleles: $p + q = 1$

B b

- Individuals: $p^2 + 2pq + q^2 = 1$

BB Bb bb

Lecture notes on

Genetic Structure of Population

First Eprint

Department of Animal Genetics & Breeding

College of Veterinary & Animal Sciences

Sardar Vallabhbhai Patel University of Agriculture & Technology

Modipuram, Meerut- 250 110

Published Online:

First eprint: 2021
Total pages: 40
2021.02.23 22:15:11 +05'30'

Published by:

Department of Animal Genetics & Breeding
College of Veterinary & Animal Sciences
Sardar Vallabhbhai Patel University of Agriculture and Technology
Meerut- 250 110, Uttar Pradesh, India

Publication No.

SVP/2021/06/02/213 Dated: February 23, 2021 *(for official use)*

How to cite this Lecture notes

Tyagi, K 2021, *Genetic Structure of Population*, lecture notes, Principles of Animal and Population Genetics AGB UNIT II, Sardar Vallabhbhai Patel University of Agriculture & Technology, Meerut, India, Delivered 19, 20 & 23 February 2021. Retrieved online from https://vepub.com

Address Correspondence to:

Dr. Kuldeep Kumar Tyagi
Associate Professor & OIC
Department of Animal Genetics & Breeding
COVAS, SVPUAT, Meerut- 250110 (U.P.) India
drtyagivet@gmail.com
+91 9601283365 (M)

ABOUT

These lecture notes on “Genetic Structure of Population” were prepared and delivered to my undergraduate students studying Animal Genetics & Breeding course. This course is offered in the second professional year of Bachelor of Veterinary Science & Animal Husbandry degree at College of Veterinary & Animal Sciences, S.V.P.U.A.T, Meerut, Uttar Pradesh, India. This lecture paves the foundation of population genetics. In this lecture the basic background of biological organization, and place of population in this biological organization have been explained. Briefing about various types of population had also been undertaken. Mendelian population and how it differed from an individual had been introduced. Concept of the gene pool and its role in the genetic structure of the population had been touched upon. An easy method of using dummy variables for the estimation of allelic and genotypic frequencies in a population had been explained. Various situations pertaining to different numbers of locus and different number of alleles for autosomes and allosomes had been explained for better in depth understanding of the topic by students. Use of explanatory illustrations, examples and figures had deliberately been used to create an interest among the students. Once through with this lecture notes readers will be able to understand the genetic structure of the population. I had tried my level best to simplify the concept in easy to understand language. Further constructive suggestions to improve this lecture notes are always welcome from readers on my email and whatsapp.

KULDEEP
KUMAR TYAGI

Digitally signed by KULDEEP KUMAR TYAGI
DN: c=IN, o=Personal, pseudonym=70794731d2ba740b74ed194a7814f376683ebedfd8ebf235a45b740007b4f233, postalCode=396450, st=Gujarat, serialNumber=539e1ff36ecbb01285636bc5c65cc1b4b14f5a1620b67a5f3749c552dce5a724, cn=KULDEEP KUMAR TYAGI
Date: 2021.02.23 22:16:15 +05'30'

DISCLAIMER

These lecture notes on "Genetic Structure of Population" have been compiled from various resources available in the public domain. Only excerpts from the original works have been used. This is being done for educational purposes in the interest of developing a concise and updated reading material for students with no intent of commercial benefits. References to the source of material used have been included in the footnotes. The author does not claim any ownership of any copyrighted material included by chance in the lecture. If due to inability to trace the original source any copyrighted material got included, it may please be brought to the notice of the author for rectification.

KULDEEP
KUMAR
TYAGI

Digitally signed by KULDEEP KUMAR TYAGI
DN: c=IN, o=Personal, pseudonym=70794731d2ba740b74ed194a7814f376683ebedfd8ebf235a45b740007b4f233, postalCode=396450, st=Gujarat, serialNumber=539e1ff36ecbb01285636bc5c65cc1b4b14f5a1620b67a5f3749c552dce5a724, cn=KULDEEP KUMAR TYAGI
Date: 2021.02.23 22:17:27 +05'30'

OTHER LECTURE NOTES BY THE AUTHOR

S.No.	Title	Date of Publication	Revisions	Access
1	Introduction and importance of statistics and biostatistics	12-10-2020	First eprint	Download
2	Probability	20-08-2020	First eprint	Download
3	Probability	24-11-2020	Second revised eprint	Download
4	Probability Distribution	07-12-2020	First eprint	Download
5	Introduction of population genetics	23-02-2021	First eprint	Download

TABLE OF CONTENTS

Genetics Structure of Population

1. Introduction

Genetic structure refers to DNA sequences as well as the various ways that combinations of alleles become packaged within individuals[1]. Population is the basic pivot on which genetic structures are deciphered in population genetics. Therefore, it becomes pertinent to understand more about population, its types and position in biological organization. In any species, a great deal of genetic variation within and between populations arises from the existence of various alleles at different gene loci. A fundamental measurement in population genetics is the frequency at which the alleles are found at any gene locus of interest[2]. Calculation of gene and genotypic frequencies in a population is made easy by the concept of gene pool. This chapter helps in understanding basic concepts, some terminologies and methodologies to calculate gene and genotypic frequencies in a population.

2. Biological organization

Biological organization is the hierarchy of complex biological structures and systems that define life using a reductionist approach[3]. Each level in the hierarchy represents an increase in organizational complexity, with each "object" being primarily composed of the previous level's basic unit[4].

Table 1 Terms used to describe biological organization

Term	Meaning
Biosphere[5,6]	The term "biosphere" was coined by geologist Eduard Suess in

[1] J.M. Herbers, 2010, Evolution: Fundamentals, Editor(s): Michael D. Breed, Janice Moore,Encyclopedia of Animal Behavior, Academic Press, Pages 670-678, ISBN 9780080453378, https://doi.org/10.1016/B978-0-08-045337-8.00110-8.

[2] Griffiths AJF, Miller JH, Suzuki DT, et al. An Introduction to Genetic Analysis. 7th edition. New York: W. H. Freeman; 2000. Chapter 24, Population Genetics. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21961/

[3] Solomon, Eldra P.; Berg, Linda R.; Martin, Diana W. (2002), Biology (6th ed.), Brooks/Cole, ISBN 0-534-39175-3, LCCN 2001095366, pp. 9–10

[4] Pavé, Alain (2006), "Biological and Ecological Systems Hierarchical organization", in Pumain, D. (ed.), Hierarchy in Natural and Social Sciences, New York, New York: Springer-Verlag, ISBN 978-1-4020-4126-6. p.40

[5] Suess, E. (1875) Die Entstehung Der Alpen [The Origin of the Alps]. Vienna: W. Braunmuller.

[6] Retrieved online https://en.wikipedia.org/wiki/Biosphere

	1875, which he defined as the place on Earth's surface where life dwells. The biosphere is the global ecological system integrating all living beings and their relationships, including their interaction with the elements of the lithosphere, geosphere, hydrosphere, and atmosphere.
Biome[7]	A biome /'baɪoʊm/ is a collection of plants and animals that have common characteristics for the environment they exist in. They can be found over a range of continents. Biomes are distinct biological communities that have formed in response to a shared physical climate.
Ecosystem[8]	An ecosystem is a community of living organisms in conjunction with the nonliving components of their environment, interacting as a system.
Community[9]	A community is a social unit (*a group of living things of different species*) with commonality such as norms, religion, values, customs, or identity. Communities may share a sense of place situated in a given geographical area (e.g. a country, village, town, or neighbourhood) or in virtual space through communication platforms.
Population[10]	In biology, a population is a number of all the organisms of the *same group or species* who live in a *particular geographical* area and are capable of interbreeding.
Species[11]	Species, in biology, classification comprising *all related organisms* that share common characteristics and are capable of interbreeding.
Organism[12]	In biology, an organism (from Greek: ὀργανισμός, organismos) is any individual entity that embodies the properties of life.

[7] Cain, Michael; Bowman, William; Hacker, Sally (2014). *Ecology* (Third ed.). Massachusetts: Sinauer. p. 51. ISBN 9780878939084

[8] Tansley (1934); Molles (1999), p. 482; Chapin et al. (2002), p. 380; Schulze et al. (2005); p. 400; Gurevitch et al. (2006), p. 522; Smith & Smith 2012, p. G-5. Retrieved online https://en.wikipedia.org/wiki/Ecosystem

[9] Revised from "Community : The Blackwell Encyclopedia of Sociology : Blackwell Encyclopedia of Social Online". *www.sociologyencyclopedia.com*.

[10] Retrieved online https://en.wikipedia.org/wiki/Population

[11] Revised from https://www.britannica.com/science/species-taxon

[12] Retrieved online https://en.wikipedia.org/wiki/Organism

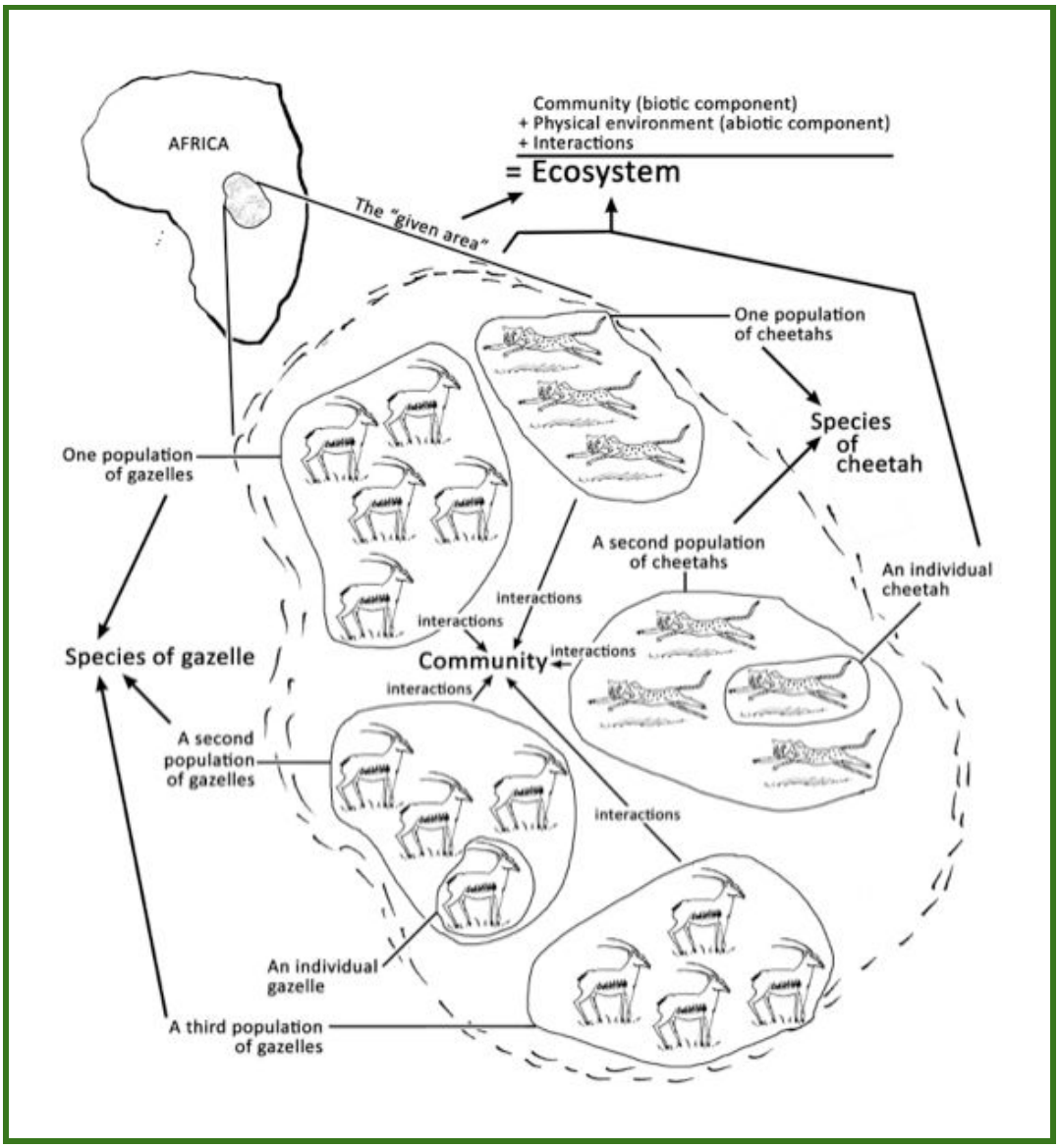

Fig 1: Interactions among different levels of biological organizations in an ecosystem[13]

3. Population

The term population had been defined differently in context to different subjects. Some of the commonly used definitions in the subject of biology, statistics and ecology have been presented in Box 1.

[13] Courtesy: Khan Academy, retrieved online from https://cdn.kastatic.org/ka-perseus-images/3139dcbcdbd8d2f70298abf639fd81a56a369903.png

Box 1: Meaning of population in context to different subjects

In biology, a population is a number of all the organisms of the same group or species who live in a particular geographical area and are capable of interbreeding.

Wikipedia[14]

In statistics, a population is the entire pool from which a statistical sample is drawn.

Will Kenton[15]

In ecology, a population consists of all the organisms of a particular species living in a given area.

Khan Academy[16]

The natural groups of individuals make up the population. They have been described and termed differently when certain preconditions are observed. Various terms used to describe natural groups of individuals in ecological studies have been presented in table 1.

Table 1 Terms used to describe natural groups of individuals in ecological studies[17]

Term	Definition
Species population	All individuals of a species.
Metapopulation	A set of spatially disjunct populations, among which there is some migration.
Population	A group of conspecific individuals that is demographically, genetically, or spatially disjunct from other groups of individuals.
Aggregation	A spatially clustered group of individuals.
Deme	A group of individuals more genetically similar to each other than to other individuals, usually with some degree of spatial isolation as well.

[14] Retrieved online https://en.wikipedia.org/wiki/Population
[15] Retrieved online https://www.investopedia.com/terms/p/population.asp
[16] Retrieved online https://www.khanacademy.org/science/high-school-biology/hs-ecology/hs-population-ecology/a/population-size-density-and-dispersal
[17] Wells, J. V.; Richmond, M. E. (1995). "Populations, metapopulations, and species populations: What are they and who should care?". *Wildlife Society Bulletin*. **23** (3): 458–462.

Local population	A group of individuals within an investigator-delimited area smaller than the geographic range of the species and often within a population (as defined above). A local population could be a disjunct population as well.
Subpopulation	An arbitrary spatially delimited subset of individuals from within a population (as defined above).

4. Mendelian Population

The population genetic approach uses the concept of the Mendelian population which Dobzhonsky has defined as a "reproductive community of sexual and cross fertilizing individuals which share in a common gene pool". The significance of Mendalian populations in evolutionary genetics lies in the web of genetic relationships within and between them—allele frequencies, consanguinity, mating patterns, gene flow, natural selection, etc. A Mendelian population, defined as a group of individuals who reproduce sexually or are potentially capable of doing so, focuses on the evolutionary dimension and differs therefore fundamentally from other disciplines like general ecology or demography[18].

Table 2 Difference between Mendalian population and individual

Mendelian population	Individual
Genes carried in a gene pool of a Mendalian population have continuity over time	Gene set of an individual does not have continuity until unless it gets cloned
Indefinite lifespan (Immortal) ie. previous population is replaced by similar newer population in absence of evolutionary forces	Individual have definite life span(mortal)
The genetic composition of a population may change due to various evolutionary forces	The genetic makeup of an individual remains intact throughout life until unless there is some mutation.

[18]Serrano Sanchez, C. Introduction: The concept of population. Int. J. Anthropol. 11, 15–18 (1996). https://doi.org/10.1007/BF02441407

5. Gene pool

The Russian geneticist *Alexander Sergeevich Serebrovsky* first formulated the concept in the 1920s as genofond (gene fund), a word that was imported to the United States from the Soviet Union by *Theodosius Dobzhansky*, who translated it into English as "gene pool"[19]. The gene pool gets its name from the idea that we are essentially taking all the gene copies—for all genes—in the individuals of a population and dumping them into one large, common pool[20]. The fact that genes exist in alternate forms, called alleles, forms the basis for the study of population genetics. Populations are made up of members of the same species that interbreed. Population geneticists study the variation that naturally occurs among the genes within a population. The collection of all the genes and the various alternate or allelic forms of those genes within a population is called its gene pool[21]. The gene pool can be defined as the set of all genes, or genetic information, in any population, usually of a particular species[22].

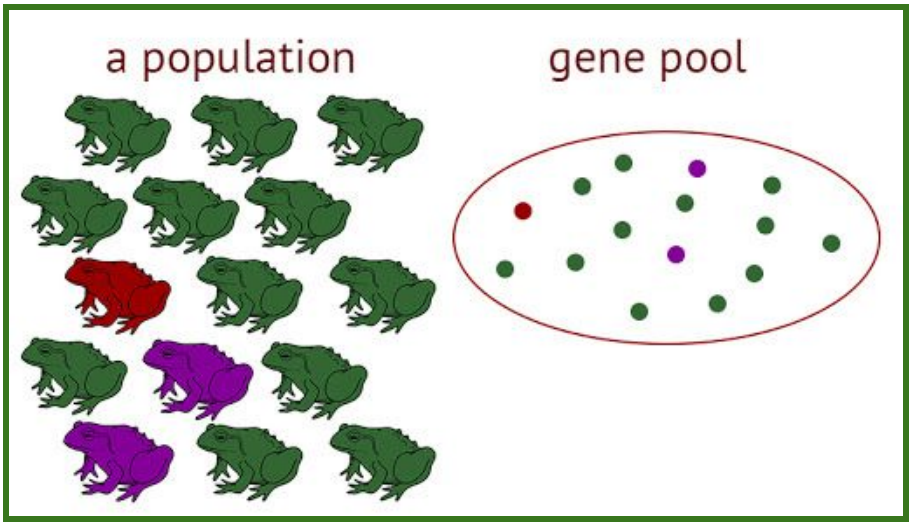

Fig 2: Illustration of gene pool[23]

[19] Graham, Loren (2013). Lonely Ideas: Can Russia Compete?. MIT Press. p. 169. ISBN 978-0-262-01979-8.
[20] Retrieved online https://www.khanacademy.org/science/ap-biology/natural-selection/hardy-weinberg-equilibrium/a/allele-frequency-the-gene-pool
[21] Retrieved online https://www.nature.com/scitable/topicpage/the-collective-set-of-alleles-in-a-6385985/
[22] Retrieved online https://www.britannica.com/science/gene-pool
[23] Retrieved online http://edelweisspublications.com/keyword/51/1971/Gene-pool

Gene pool indicates extensive genetic diversity found within a population. This is associated with strong populations that can survive bouts of intense selection. When all individuals in a population are identical with regard to a particular phenotypic trait and the population is known to be monomorphic. When the individuals show differences in a particular trait they are known as polymorphic. Most populations have some degree of variation in their gene pools. By measuring the amount of genetic variation in a population, scientists can begin to make predictions about how genetic variation changes over time. These predictions can then help them gain important insights into the processes that allow organisms to adapt to their environment or to develop into new species over generations, also known as the process of evolution[24]. Composition of a population gene pool can be changed over time through evolution. This can occur by a variety of mechanisms which includes mutations, genetic drift and natural selection.

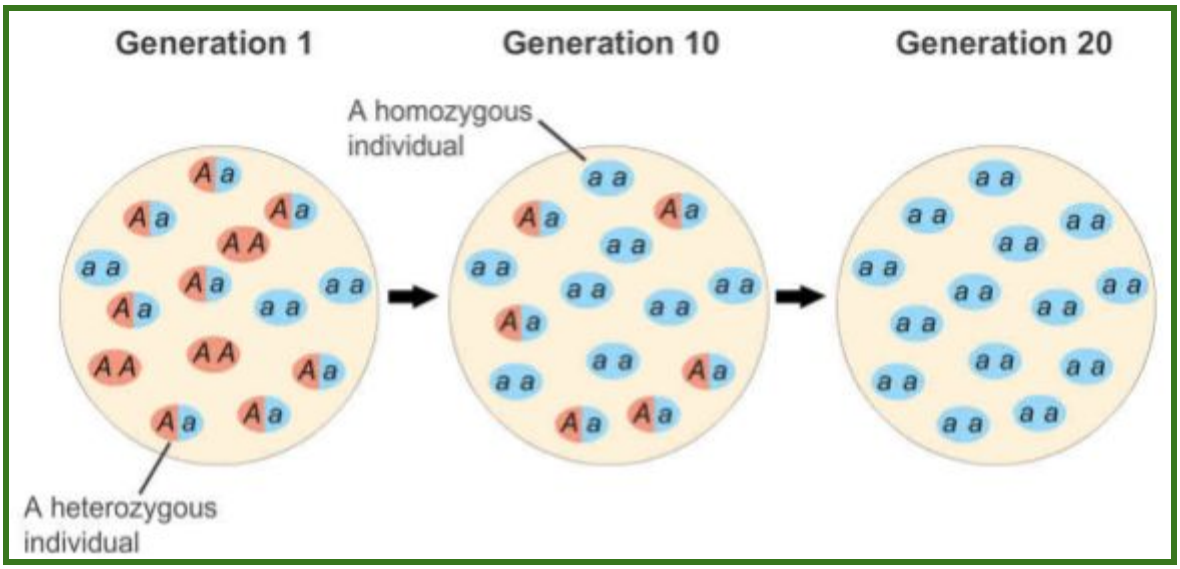

Fig 3: Gene pool of a population over generations[25]

5.1. Classification of population based on gene pool

Harlan and de Wet (1971) proposed classifying each crop and its related species by gene pools rather than by formal taxonomy[26].

[24] Retrieved online https://www.nature.com/scitable/topicpage/the-variety-of-genes-in-the-gene-6526291/

[25] Retrieved online https://www.whfreeman.com/BrainHoney/Resource/6716/SitebuilderUploads/Hillis2e/Student%20Resources/Animated%20Tutorials/pol2e_at_1502_genetic_drift_simulation/pol2e_at_1502_genetic_drift_simulation.html

[26] Harlan, J.R.; Wet, J.M.J.d. (1971). "Toward a Rational Classification of Cultivated Plants". *Taxon*. **20** (4): 509–517. doi:10.2307/1218252

5.1.1. Primary gene pool

Members of this gene pool are probably in the same "species" (in conventional biological usage) and can intermate freely. Harlan and de Wet wrote, "Among forms of this gene pool, crossing is easy; hybrids are generally fertile with good chromosome pairing; gene segregation is approximately normal and gene transfer is generally easy." They also advised subdividing each crop gene pool in two:

- Subspecies A: Cultivated races
- Subspecies B: Spontaneous races (wild or weedy)

5.1.2. Secondary gene pool

Members of this pool are probably normally classified as different species than the crop species under consideration (the primary gene pool). However, these species are closely related and can cross and produce at least some fertile hybrids. As would be expected by members of different species, there are some reproductive barriers between members of the primary and secondary gene pools:

- hybrids may be weak
- hybrids may be partially sterile
- chromosomes may pair poorly or not at all
- recovery of desired phenotypes may be difficult in subsequent generations
- However, "The gene pool is available to be utilized, however, if the plant breeder or geneticist is willing to put out the effort required.

5.1.3. Tertiary gene pool

Members of this gene pool are more distantly related to the members of the primary gene pool. The primary and tertiary gene pools can be intermated, but gene transfer between them is impossible without the use of "rather extreme or radical measures" such as:

- embryo rescue (or embryo culture, a form of plant organ culture)
- induced polyploidy (chromosome doubling)

- bridging crosses (e.g., with members of the secondary gene pool).

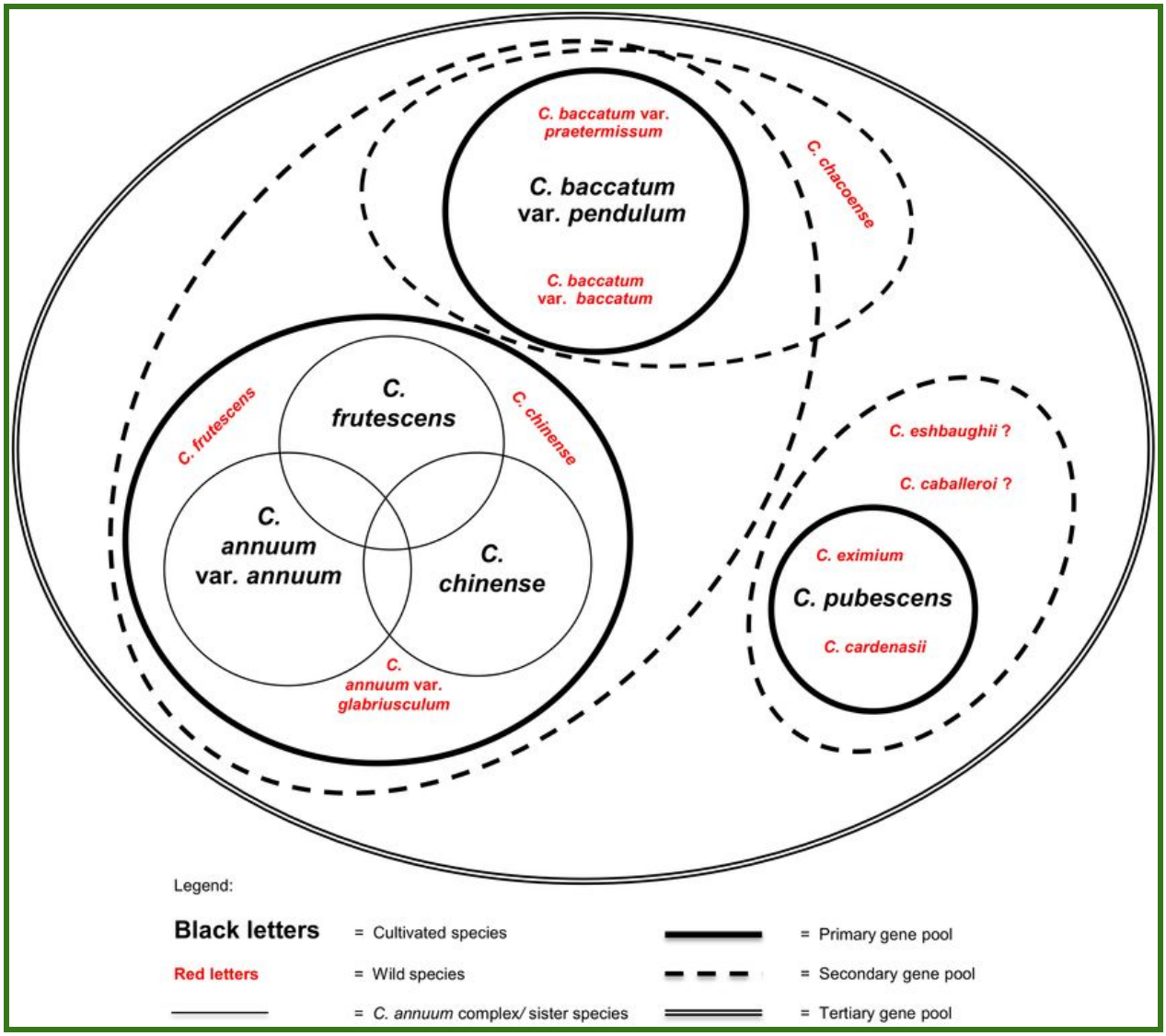

Fig 4: Gene pools of cultivated capsicum[27]

5.2. Gene pool centres

Gene pool centres refers to areas on the earth where important crop plants and domestic animals originated. They have an extraordinary range of the wild counterparts of cultivated plant species and useful tropical plants. Gene pool centres also contain

[27] van Zonneveld, et al. (2015). Screening Genetic Resources of Capsicum Peppers in Their Primary Center of Diversity in Bolivia and Peru. PloS one. 10. e0134663. 10.1371/journal.pone.0134663.

different sub tropical and temperate region species.

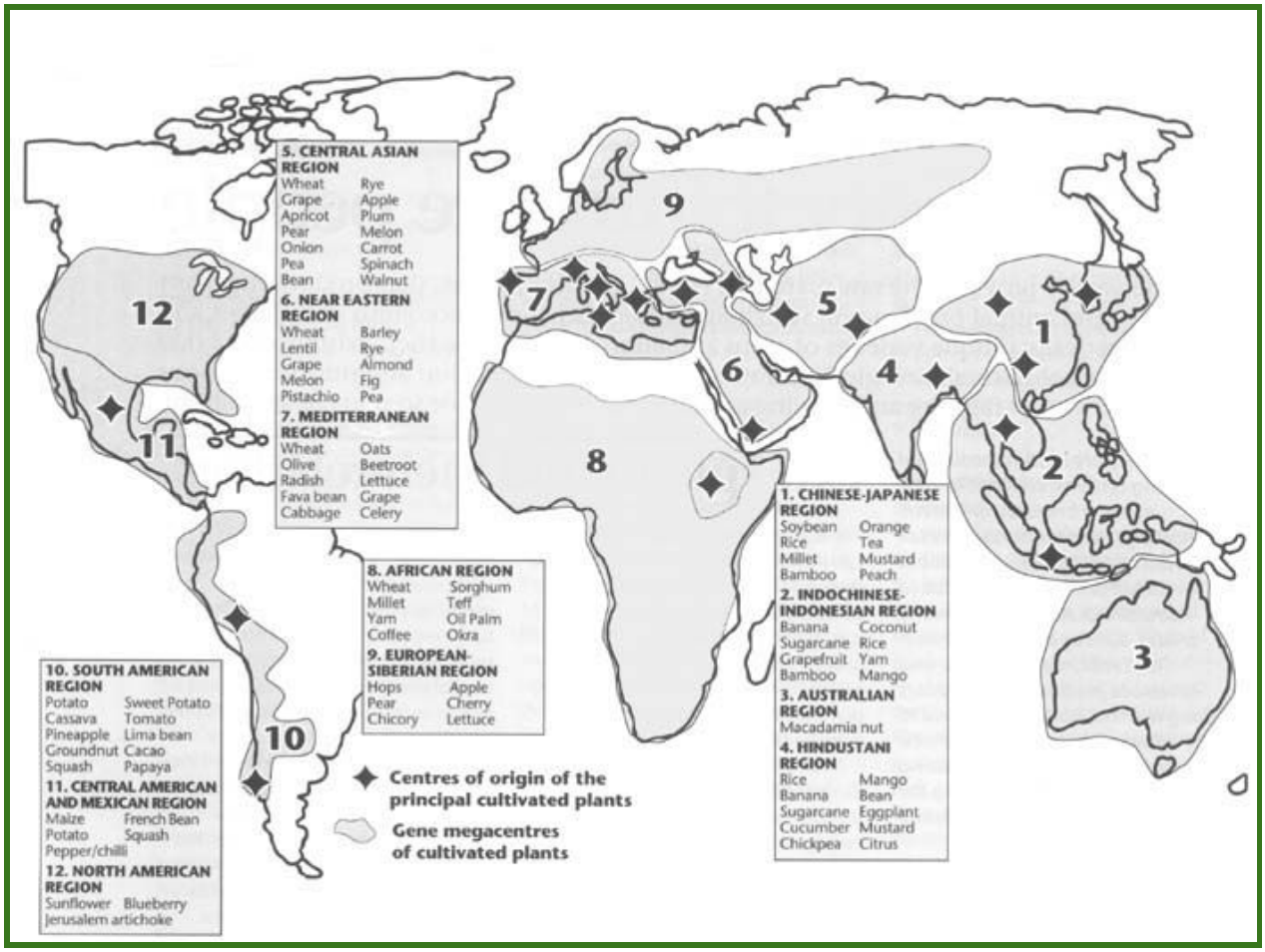

Fig 5: The twelve megacentres of cultivated plants[28]

6. Population genetic structure

Genetic structure refers to any pattern in the genetic makeup of individuals within a population[29]. Genetic structure refers to DNA sequences as well as the various ways that combinations of alleles become packaged within individuals. Thus, a population's genetic structure includes the component of genotype frequencies within the genotype pool as well as allele frequencies in the gene pool, combinations of alleles at different

[28] Retrieved online http://www.fao.org/3/V1430E/V1430E04.htm

[29] Chakraborty, Ranajit (1993). "Analysis of Genetic Structure of Populations: Meaning, Methods, and Implications". *Human Population Genetics*. Springer, Boston, MA. pp. 189–206. doi:10.1007/978-1-4615-2970-5_14

genetic loci, and even chromosomal rearrangements[30].

We will take up the first two components i) genotype frequencies within the genotype pool and ii) allele frequencies in the gene pool, to understand the concept of population genetic structure in an easy to understand manner. We will be concentrating on diploid organisms since all or nearly all mammals are diploid organisms. We will use a simple tabular method to calculate gene and genotypic frequencies under various modes of inheritance at single and multiple locus.

6.1. Single autosomal locus

An autosome is any chromosome that is not a sex chromosome[31]. In genetics, a locus (plural loci) is a specific, fixed position on a chromosome where a particular gene or genetic marker is located[32]. Genes may possess multiple variants known as alleles, and an allele may also be said to reside at a particular locus. A gene is said to be polymorphic if more than one allele occupies that gene's locus within a population[33]. In addition to having more than one allele at a specific locus, each allele must also occur in the population at a rate of at least 1% to generally be considered polymorphic[34]. However, what is important at a locus is the degree of polymorphism, and a locus in which there are 1,000 equi frequent alleles would be considered much more polymorphic than a locus at which there are two alleles with frequencies 0.01 and 0.99. Many authors now use the term mutation for any rare allele, and the term polymorphism for any common allele[35]. Now let us consider some of the examples under various modes of inheritance at a single polymorphic locus.

6.1.1. Two alleles

The genetic structure of population under various modes of inheritance have been given below with the help of illustrations and examples.

---

[30] J.M. Herbers, in Encyclopedia of Animal Behavior, 2010

[31] Griffiths, Anthony J. F. (1999). An Introduction to genetic analysis. New York: W.H. Freeman. ISBN 978-0-7167-3771-1.

[32] Wood, E.J. (1995). "The encyclopedia of molecular biology". Biochemical Education. 23 (2): 1165. doi:10.1016/0307-4412(95)90659-2

[33] Genetic polymorphism - Biology-Online Dictionary | Biology-Online Dictionary

[34] Ford EB. Polymorphism and taxonomy. In: Huxley J, editor. The new systematics. Oxford: 1940.

[35] Elston, R. C., Satagopan, J. M., & Sun, S. (2012). Genetic terminology. Methods in molecular biology (Clifton, N.J.), 850, 1–9. https://doi.org/10.1007/978-1-61779-555-8_1

Illustration 1: Beta-lactoglobulin in milk of cows[36]

The gene underlying the production of beta-lactoglobulin protein in cow's milk shows polymorphisms which affect the amount of protein produced. Higher levels of beta-lactoglobulin are associated with increased milk yield and whey protein content, and low levels are associated with increased casein and fat content and are favorable for cheese production.

Variants:

- A = associated with higher relative content of beta-lactoglobulin
- B = associated with lower relative content of beta-lactoglobulin

Applicable to: Many dairy breeds

Mode of Inheritance: Incomplete dominance

Phenotypic effect:

- Cows with A/A genotype will produce higher amounts of beta-lactoglobulin.
- Cows with A/B genotype will produce intermediate amounts of beta-lactoglobulin.
- Cows with B/B genotype will produce low amounts of beta-lactoglobulin.

Example 1: Three hundred holstein friesian cows were tested for Beta-lactoglobulin gene polymorphism at SNP, g.-731G>A. The three genotyped forms obtained were as follows

AA	AB	BB
90	150	60

Calculate the gene and genotypic frequency of this sample from the population?

Solution 1:
Since the mode of inheritance was incomplete dominance, the three genotypes can very well be differentiated phenotypically on the basis of Beta-lactoglobulin amount

[36] Illustration adapted from Ganai, N.A., Bovenhuis, H., van Arendonk, J.A., & Visker, M.H. (2009). Novel polymorphisms in the bovine beta-lactoglobulin gene and their effects on beta-lactoglobulin protein concentration in milk. Animal Genetics, 40(2), 127-133. doi: 10.1111/j.1365-2052.2008.01806.x

in the milk.

A	B	Genotypes	Number (n)	n × A	n × B
2	0	AA	90	$2 \times 90 = 180$	$0 \times 90 = 0$
1	1	AB	150	$1 \times 150 = 150$	$1 \times 150 = 150$
0	2	BB	60	$0 \times 60 = 0$	$2 \times 60 = 120$
		Sum	$N_{..} = 300$	$n_A = 330$	$n_B = 270$

Here, two alleles A and B are two dummy variables that can take any dummy value among 0, 1 and 2 such that A + B will always be equal to two (2). Since a diploid individual can only have 2 alleles at a time.

Genotypic frequencies

Now, genotypic frequency f(..) is the relative frequency of that genotype in the genotype pool and can alternatively be termed as probability of occurrence of a particular genotype in the genotype pool.

$$p(AA) = \frac{n_{AA}}{N_{..}} \Rightarrow \frac{90}{300} = 0.3$$

$$p(AB) = \frac{n_{AB}}{N_{..}} \Rightarrow \frac{150}{300} = 0.5$$

$$p(BB) = \frac{n_{BB}}{N_{..}} \Rightarrow \frac{60}{300} = 0.2$$

$$\text{and } p(..) = p(AA) + p(AB) + p(BB) = 1$$

Thus the formula for calculating genotypic frequencies in a diploid mendelian population for a locus with two alleles can be generalized as

$$p(..) = \frac{n_{..}}{N_{..}}$$

Where, $N_{..} = \sum n_{..}$ is the sum total of all the genotypes in the genotype pool.

$n_{..}$ is the number of individuals with a particular genotype in the genotype pool.

Here we should note that $\sum p(..) = 1$,

Therefore, function p(..) qualifies both the below given conditions to be called as probability mass function,

1. Probability of an event cannot be negative, i.e., for any value of .., $p(..) \geq 0$
2. Probabilities of all possible outcomes sum to unity, i.e., $\sum_{all\ ..} p(..) = 1$

Allelic frequencies

Now, from the table sum total of all the alleles in gene pool can be calculated as

$$N_{.} = \sum_{all\ .} n_{.} = n_A + n_B \Rightarrow 330 + 270 = 600$$

Now, allelic frequency f(.) is the relative frequency of that allele in the gene pool and can alternatively be termed as probability of occurrence of a particular allele in the gene pool.

$$p(A) = \frac{n_A}{N_{.}} \Rightarrow \frac{330}{600} = 0.55$$

$$p(B) = \frac{n_B}{N_{.}} \Rightarrow \frac{270}{600} = 0.45$$

$$\text{and } p(.) = p(A) + p(B) = 1$$

Thus the formula for calculating allelic frequencies in a diploid mendelian population for a locus with two alleles can be generalized as

$$p(.) = \frac{n_{.}}{N_{.}}$$

Where,

$N_{.} = \sum n_{.}$ is the sum total of all the alleles in the gene pool.

$n_{.}$ is the number of particular alleles in the gene pool.

Here we should note that $\sum p(.) = 1$,

Therefore, function p(.) qualifies both the below given conditions to be called as probability mass function,

1. Probability of an event cannot be negative, i.e., for any value of ., $p(.) \geq 0$
2. Probabilities of all possible outcomes sum to unity, i.e., $\sum_{all\ .} p(.) = 1$

6.1.2. Three alleles

The genetic structure of population for a trait with three alleles under various modes of inheritance have been given below with the help of illustrations and examples.

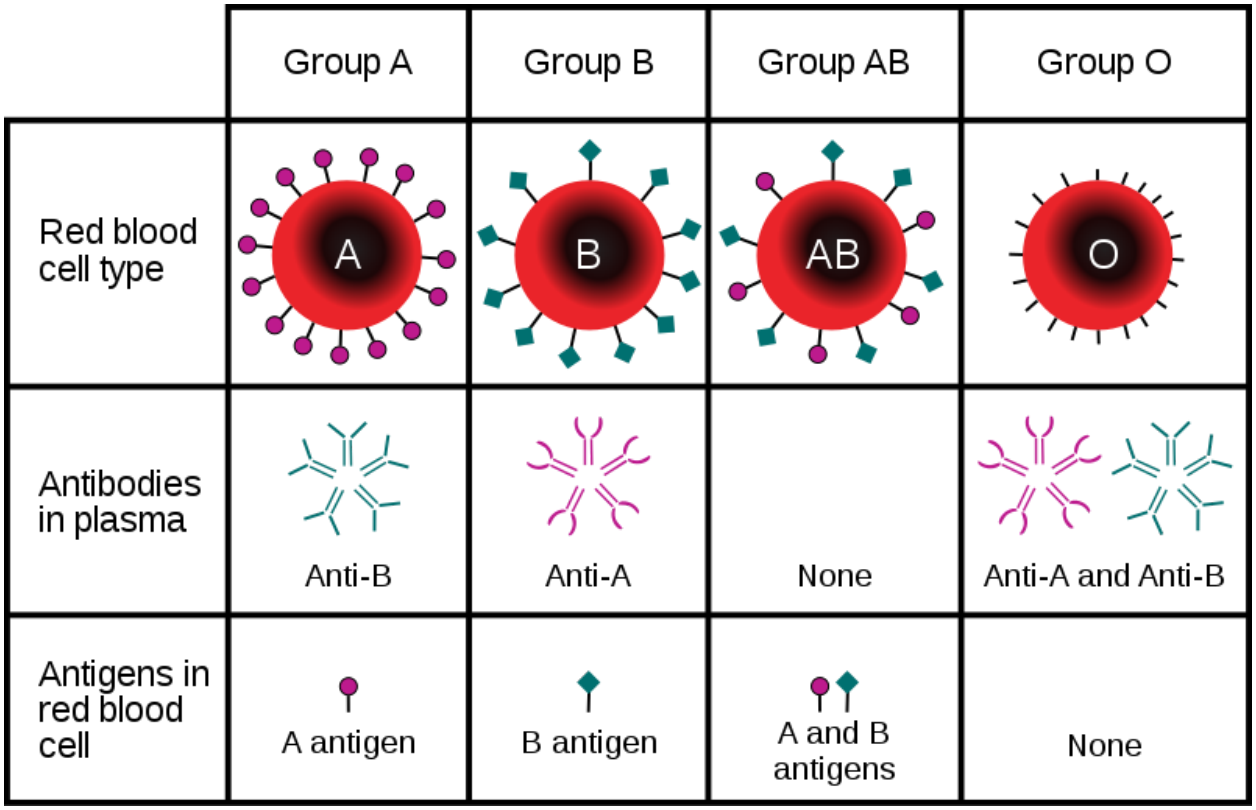

Fig 6: ABO blood types [37]

Illustration 2: ABO blood group system in humans

The ABO blood types were discovered by Karl Landsteiner in 1901; he received the

[37] Retrieved online https://commons.wikimedia.org/wiki/File:ABO_blood_type.svg

Nobel Prize in Physiology or Medicine in 1930 for this discovery[38]. The ABO blood group system is used to denote the presence of one, both, or neither of the A and B antigens on erythrocytes[39]. The ABO blood type is controlled by a single gene (the ABO gene) with three types of alleles inferred from classical genetics: i, I^A, and I^B. We will use notations A, B, O for sake of easy to understand statistical denominations.

Variants:

- A = Codominant over B but dominant over O
- B = Codominant over A but dominant over O
- O = Recessive to both A and B

Applicable to: Humans, other primates such as apes and Old World monkeys.

Mode of Inheritance: Both autosomal codominant and recessive

Phenotypic effect:

- Individuals with AA, AO genotype will have A blood group.
- Individuals with BB, BO genotype will have B blood group.
- Individuals with AB genotype will have AB blood group.
- Individuals with OO genotype will have O blood group.

Example 2: A sample human population of 900 individuals were screened genetically for the presence of ABO alleles. Following genotypes were mapped to different individuals in the sample

AA	AO	BB	BO	AB	OO
150	90	120	180	300	60

Calculate the gene and genotypic frequency of this sample from the population?

Solution 2:

The tabulation for deducing gene and genotypic frequencies will be as follows

[38] Maton, Anthea; Jean Hopkins; Charles William McLaughlin; Susan Johnson; Maryanna Quon Warner; David LaHart; Jill D. Wright (1993). Human Biology and Health. Englewood Cliffs, New Jersey, USA: Prentice Hall. ISBN 978-0-13-981176-0.

[39] The Editors of Encyclopædia Britannica. "ABO blood group system". Encyclopædia Britannica. Encyclopædia Britannica, Inc.

A	B	O	Genotypes	Phenotypes	n	n × A	n × B	n × O
2	0	0	AA	A	150	2 × 150 =300	0 × 150 =0	0 × 150 =0
1	0	1	AO	A	90	1 × 90 =90	0 × 90 =0	1 × 90 =90
0	2	0	BB	B	120	0 × 120 =0	2 × 120 =240	0 × 120 =0
0	1	1	BO	B	180	0 × 180 =0	1 × 180 =180	1 × 180 =180
1	1	0	AB	AB	300	1 × 300 =300	1 × 300 =300	0 × 300 =0
0	0	2	OO	O	60	0 × 60 =0	0 × 60 =0	2 × 60 =120
			Sum		$N_{..}$ =900	n_A =690	n_B =720	n_O =390

Here, three alleles A, B and O can be taken as three dummy variables that can take any dummy value among 0, 1 and 2 such that A + B + O will always be equal to two (2). Since a diploid individual can only have 2 alleles at a time.

Genotypic frequencies

Now, genotypic frequency f(..) is the relative frequency of that genotype in the genotype pool and can alternatively be termed as probability of occurrence of a particular genotype in the genotype pool.

Formula for calculating genotypic frequencies in a diploid mendelian population for a locus with two alleles can be generalized as

$$p(..) = \frac{n_{..}}{N_{..}}$$

Where, $N_{..} = \sum n_{..}$ is the sum total of all the genotypes in the genotype pool.

$n_{..}$ is the number of individuals with a particular genotype in the genotype pool.
Provided, function p(..) qualifies both the below given conditions to be called as probability mass function,

1. Probability of an event cannot be negative, i.e., for any value of .., $p(..) \geq 0$
2. Probabilities of all possible outcomes sum to unity, i.e., $\sum_{all\ ..} p(..) = 1$

$$p(AA) = \frac{n_{AA}}{N_{..}} \Rightarrow \frac{150}{900} = 0.167$$

$$p(AO) = \frac{n_{AO}}{N_{..}} \Rightarrow \frac{90}{900} = 0.100$$

$$p(BB) = \frac{n_{BB}}{N_{..}} \Rightarrow \frac{120}{900} = 0.133$$

$$p(BO) = \frac{n_{BO}}{N_{..}} \Rightarrow \frac{180}{900} = 0.200$$

$$p(AB) = \frac{n_{AB}}{N_{..}} \Rightarrow \frac{300}{900} = 0.330$$

$$p(OO) = \frac{n_{OO}}{N_{..}} \Rightarrow \frac{60}{900} = 0.067$$

and $p(..) = p(AA) + p(AO) + p(BB) + p(BO) + p(AB) + p(OO) = 1$

Allelic frequencies

Now, from the table sum total of all the alleles in gene pool can be calculated as

$$N_{.} = \sum_{all\ .} n_{.} = n_A + n_B + n_O \Rightarrow 690 + 720 + 390 = 1800$$

Now, allelic frequency f(.) is the relative frequency of that allele in the gene pool and can alternatively be termed as probability of occurrence of a particular allele in the gene pool.

Formula for calculating allelic frequencies in a diploid mendelian population for a locus with three alleles can be generalized as

$$p(.) = \frac{n_{.}}{N_{.}}$$

Where,

$N_{.} = \sum n_{.}$ is the sum total of all the alleles in the gene pool.

$n_{.}$ is the number of particular alleles in the gene pool.

Provided, function p(.) qualifies both the below given conditions to be called as probability mass function,

1. Probability of an event cannot be negative, i.e., for any value of ., $p(.) \geq 0$
2. Probabilities of all possible outcomes sum to unity, i.e., $\sum_{all\ .} p(.) = 1$

$$p(A) = \frac{n_A}{N_.} \Rightarrow \frac{690}{1800} = 0.383$$

$$p(B) = \frac{n_B}{N_.} \Rightarrow \frac{720}{1800} = 0.400$$

$$p(O) = \frac{n_O}{N_.} \Rightarrow \frac{390}{1800} = 0.217$$

$$\text{and } p(.) = p(A) + p(B) + p(O) = 1$$

****

6.1.3. k alleles

The genetic structure of a population for a trait with "k" alleles under various modes of inheritance can be deduced to a generalized formula.

The tabulation for deducing gene and genotypic frequencies for a single locus with k alleles (A1,A2, ...,Ak) will be as follows.

S.No.	**A1**	**A2**	**.**	**Ak**	**Genotypes**	**n**	**n × A1**	**n × A2**	**n × .**	**n × Ak**
1	2	0	0	0	A1A1	n_{A1A1}	$2 \times n_{A1A1}$	$0 \times n_{A1A1}$	$0 \times n_{A1A1}$	$0 \times n_{A1A1}$
2	1	1	0	0	A1A2	n_{A1A2}	$1 \times n_{A1A2}$	$1 \times n_{A1A2}$	$0 \times n_{A1A2}$	$0 \times n_{A1A2}$
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
$\frac{k(k+1)}{2}$	0	0	0	2	AkAk	n_{AkAk}	$0 \times n_{AkAk}$	$0 \times n_{AkAk}$	$0 \times n_{AkAk}$	$2 \times n_{AkAk}$
					Sum	$N_{..}$	n_{A1}	n_{A2}	$n_.$	n_{Ak}

If there are k alleles, then number of possible genotypes in the population will be

$$= \frac{k(k+1)}{2}$$

ie. if we have k =2, Number of genotypes = $\frac{2(2+1)}{2} = 3$

if we have k =3, Number of genotypes = $\frac{3(3+1)}{2} = 6$

Here, k alleles A1, A2 ..... Ak can be taken as k dummy variables that can take any dummy value among 0, 1 and 2 such that in each row A1 + A2 + ...+ Ak will always be equal to two (2). Since a diploid individual at a time can only have 2 alleles out of all possible alleles.

Genotypic frequencies

Here, sum total of all possible genotypes in genotype pool will be $N_{..}$ and,

$$N_{..} = n_{A1A1} + n_{A1A2} + \ldots\ldots\ldots\ldots\ldots\ldots + n_{AkAk}$$

Now, genotypic frequency f(..) is the relative frequency of that genotype in the genotype pool and can alternatively be termed as probability of occurrence of a particular genotype in the genotype pool.

Formula for calculating genotypic frequencies in a diploid mendelian population for a locus with k alleles can be generalized as

$$p(..) = \frac{n_{..}}{N_{..}}$$

Where, $N_{..} = \sum n_{..}$ is the sum total of all the genotypes in the genotype pool.

$n_{..}$ is the number of individuals with a particular genotype in the genotype pool.
Provided, function p(..) qualifies both the below given conditions to be called as probability mass function,

1. Probability of an event cannot be negative, i.e., for any value of .., $p(..) \geq 0$
2. Probabilities of all possible outcomes sum to unity, i.e., $\sum_{all\ ..} p(..) = 1$

Therefore,

$$p(A1A1) = \frac{n_{A1A1}}{N_{..}}$$

$$p(A1A2) = \frac{n_{A1A2}}{N_{..}}$$

.

.

$$p(AkAk) = \frac{n_{AkAk}}{N_{..}}$$

$$\text{and } p(..) = p(A1A1) + p(A1A2) + \ldots\ldots\ldots\ldots + p(AkAk) = 1$$

Allelic frequencies

Now, from the table sum total of all the k alleles in gene pool can be calculated as

$$N_{.} = \sum_{all\ .} n_{.} = n_{A1} + n_{A2} + \ldots\ldots\ldots\ldots\ldots + n_{Ak}$$

Now, allelic frequency f(.) is the relative frequency of that allele in the gene pool and can alternatively be termed as probability of occurrence of a particular allele in the gene pool.

Formula for calculating allelic frequencies in a diploid mendelian population for a locus with three alleles can be generalized as

$$p(.) = \frac{n_{.}}{N_{.}}$$

Where,

$N_{.} = \sum n_{.}$ is the sum total of all the alleles in the gene pool.

$n_{.}$ is the number of particular alleles in the gene pool.

Provided, function p(.) qualifies both the below given conditions to be called as probability mass function,

1. Probability of an event cannot be negative, i.e., for any value of ., $p(.) \geq 0$
2. Probabilities of all possible outcomes sum to unity, i.e., $\sum_{all\ .} p(.) = 1$

$$p(A1) = \frac{n_{A1}}{N_{.}}$$

$$p(A2) = \frac{n_{A2}}{N_.}$$

.

.

$$p(Ak) = \frac{n_{Ak}}{N_.}$$

$$\text{and } p(.) = p(A1) + p(A2) + \ldots\ldots\ldots\ldots\ldots\ldots + p(Ak) = 1$$

6.2. Single allosomal locus

An allosome is a sex chromosome that differs from an ordinary autosome in form, size, and behavior. All diploid organisms with allosome-determined sex get half of their allosomes from each of their parents. In mammals, females are XX, they can pass along either of their X's, and since the males are XY they can pass along either an X or a Y. For a mammal to be female, the individual must receive an X chromosome from both parents, whereas to be male, the individual must receive a X chromosome from their mother and a Y chromosome from their father. In humans, X chromosome carries about 1500 genes, more than any other chromosome in the human body. Most of them code for something other than female anatomical traits. The Y chromosome carries about 78 genes. Most of the Y chromosome genes are involved with essential cell house-keeping activities and sperm production. Only one of the Y chromosome genes, the SRY gene, is responsible for male anatomical traits. Recombination between the X and Y chromosomes proved harmful—it resulted in males without necessary genes formerly found on the X chromosome, and females with unnecessary or even harmful genes previously only found on the Y chromosome. As a result, genes beneficial to males accumulated near the sex-determining genes, and recombination in this region was suppressed in order to preserve this male specific region[40]. Over time, the Y chromosome changed in such a way as to inhibit the areas around the sex determining genes from recombining at all with the X chromosome. As a result of this process, 95% of the human Y chromosome is unable to recombine and is called the non-recombining region of the Y chromosome (NRY) or male-specific region (MSY). Only the tips of the

[40] Graves JA (March 2006). "Sex chromosome specialization and degeneration in mammals". Cell. 124 (5): 901–14. doi:10.1016/j.cell.2006.02.024

Y and X chromosomes recombine. The tips of the Y chromosome that could recombine with the X chromosome are referred to as the pseudoautosomal region (PAR).

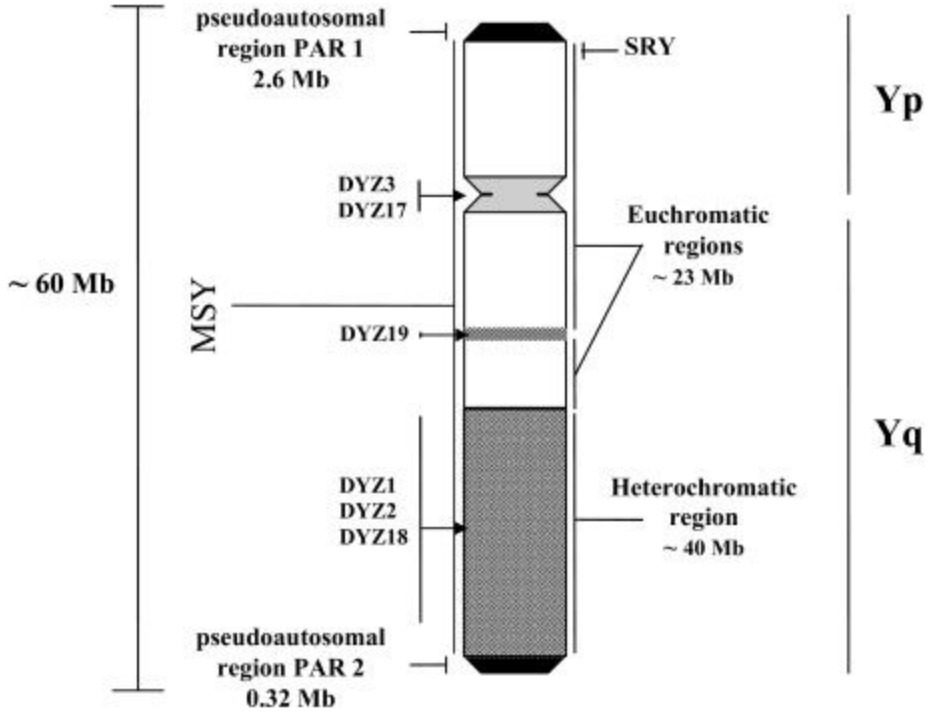

Fig 7. Y Chromosome structure[41]

6.2.1. Two alleles

The genetic structure of population for a trait governed by sex linked gene having two have been given below with the help of illustrations and examples.

Illustration 3: X-Linked Hypohidrotic Ectodermal Dysplasia in dogs[42]

X-linked hypohidrotic ectodermal dysplasia (XLHED) is a genetic disorder characterized by abnormalities in ectodermal derivatives such as sweat glands, hair, and teeth. In animals, the highest number of cases has been reported in dogs, which show characteristic congenital alopecia and develop abnormalities in the shape and number of teeth.

[41] Leonor GusmãoMaría BriónIva Gomes, in Handbook of Analytical Separations, 2008

[42] Enio Moura et al. 2019. X-Linked Hypohidrotic Ectodermal Dysplasia—General Features and Dental Abnormalities in Affected Dogs Compared With Human Dental Abnormalities, Topics in Companion Animal Medicine,Volume 5, Pages 11-17, ISSN 1938-9736, https://doi.org/10.1053/j.tcam.2019.03.002.

Mode of Inheritance: X-linked recessive

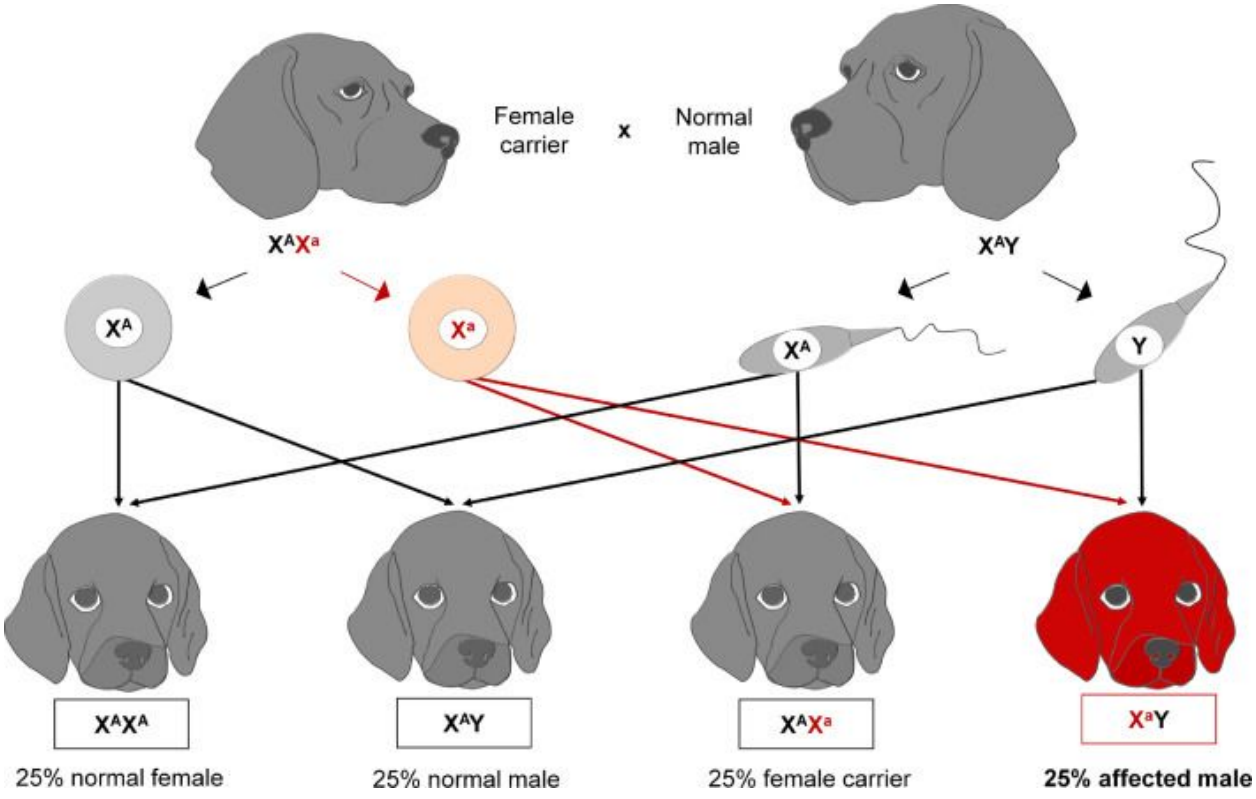

Fig 8. Probability of being born with XLHED from the mating between a female carrier and a normal male[40].

Variants: Males only have one X chromosome whereas females have two, therefore possible genotypes will differ by sex.

- Female dogs with A/A genotype and male dogs with A genotype will not have XLHED, and cannot transmit this variant to their offspring.
- Female dogs with A/a genotype will not have XLHED, but are carriers. If a carrier female is bred to a normal male, all female puppies will be normal but 50% of them will be carriers. Among male puppies from this type of cross, 50% will be normal and 50% will be affected by XLHED.
- Female dogs with a/a genotype and male dogs with “a” genotype will have XLHED.

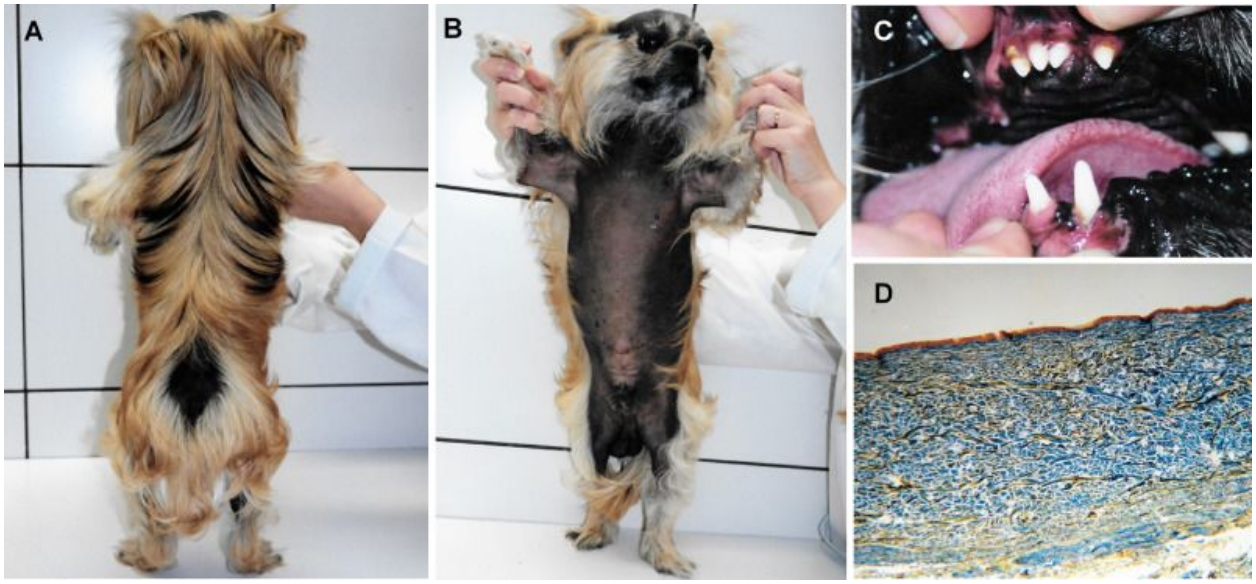

Fig 8. Clinical phenotype. (A, B) Pekingese crossbreed at 9 years of age affected by XLHED, showing the characteristic alopecic regions. (C) Oligodontia and conoid teeth. Note that there is mesioversion of 3 of the upper incisors and labioversion of the lower incisors. (D) Histologic section of an area of alopecia showing hyperkeratosis, normal collagenous fibers, and absence of piloglandular units (Mallory, 40×)[43].

Example 2: A sample Pekingese crossbreed dog population of 1000 dogs was screened genetically for the presence of X-Linked Hypohidrotic Ectodermal Dysplasia alleles. Following genotypes were mapped to different dogs in the sample

Female dogs			Male dogs	
AA	Aa	aa	A	a
200	250	50	400	100

Calculate the gene and genotypic frequency of this sample from the population?

Solution 3:

The gene and genotypic frequencies in case of sex linked traits are calculated separately for males and females. Here, females act as a diploid whereas males act as haploid organisms.

[43] Enio Moura et al. 2019. X-Linked Hypohidrotic Ectodermal Dysplasia—General Features and Dental Abnormalities in Affected Dogs Compared With Human Dental Abnormalities, Topics in Companion Animal Medicine,Volume 5, Pages 11-17, ISSN 1938-9736, https://doi.org/10.1053/j.tcam.2019.03.002.

Female population:

The tabulation for deducing gene and genotypic frequencies for females will be the same as described in example 1. Let us assume that allele "A" will be denoted as "A" and allele "a" will be denoted as "B" for sake of easy to understand statistical denominations.

A	B	Genotypes	Number (n)	n × A	n × B
2	0	AA	200	$2 \times 200 = 400$	$0 \times 200 = 0$
1	1	AB	250	$1 \times 250 = 250$	$1 \times 250 = 250$
0	2	BB	50	$0 \times 50 = 0$	$2 \times 50 = 100$
		Sum	$N_{..} = 500$	$n_A = 650$	$n_B = 350$

Here, two alleles A and B are two dummy variables that can take any dummy value among 0, 1 and 2 such that A + B will always be equal to two (2). Since a diploid individual can only have 2 alleles at a time.

Genotypic frequencies

Now, genotypic frequency f(..) is the relative frequency of that genotype in the genotype pool and can alternatively be termed as probability of occurrence of a particular genotype in the genotype pool.

Formula for calculating genotypic frequencies in a diploid mendelian population for a locus with two alleles can be generalized as

$$p(..) = \frac{n_{..}}{N_{..}}$$

Where, $N_{..} = \sum n_{..}$ is the sum total of all the genotypes in the genotype pool.

$n_{..}$ is the number of individuals with a particular genotype in the genotype pool.
Provided, function p(..) qualifies both the below given conditions to be called as probability mass function,

1. Probability of an event cannot be negative, i.e., for any value of .., $p(..) \geq 0$

2. Probabilities of all possible outcomes sum to unity, i.e., $\sum_{all\ ..} p(..) = 1$

$$p(AA) = \frac{n_{AA}}{N_{..}} \Rightarrow \frac{200}{500} = 0.4$$

$$p(AB) = \frac{n_{AB}}{N_{..}} \Rightarrow \frac{250}{500} = 0.5$$

$$p(BB) = \frac{n_{BB}}{N_{..}} \Rightarrow \frac{50}{500} = 0.1$$

$$\text{and } p(..) = p(AA) + p(AB) + p(BB) = 1$$

Allelic frequencies

Now, from the table, sum total of all the alleles in gene pool can be calculated as

$$N_{.} = \sum_{all\ .} n_{.} = n_A + n_B \Rightarrow 650 + 350 = 1000$$

Now, allelic frequency f(.) is the relative frequency of that allele in the gene pool and can alternatively be termed as probability of occurrence of a particular allele in the gene pool.

Formula for calculating allelic frequencies in a diploid mendelian population for a locus with two alleles can be generalized as

$$p(.) = \frac{n_{.}}{N_{.}}$$

Where,

$N_{.} = \sum n_{.}$ is the sum total of all the alleles in the gene pool.

$n_{.}$ is the number of particular alleles in the gene pool.

Provided, function p(.) qualifies both the below given conditions to be called as probability mass function,

1. Probability of an event cannot be negative, i.e., for any value of ., $p(.) \geq 0$
2. Probabilities of all possible outcomes sum to unity, i.e., $\sum_{all\ .} p(.) = 1$

$$p(A) = \frac{n_A}{N_.} \Rightarrow \frac{650}{1000} = 0.65$$

$$p(B) = \frac{n_B}{N_.} \Rightarrow \frac{350}{1000} = 0.35$$

$$\text{and } p(.) = p(A) + p(B) = 1$$

Male population:

The tabulation for deducing gene and genotypic frequencies for males will be somewhat different as males will act as haploid organisms for X linked traits. Let us assume that allele "A" will be denoted as "A" and allele "a" will be denoted as "B" for sake of easy to understand statistical denominations.

A	B	Genotypes	Number (n)	n × A	n × B
1	0	A-	400	1 × 400 =400	0 × 200 =0
0	1	B-	100	0 × 100 =0	1 × 100 =100
		Sum	$N_{.-}$ =500	n_A =400	n_B =100

Here, two alleles A and B are two dummy variables that can take any dummy value among 0 and 1 such that A + B will always be equal to one (1). Since male individuals will act as a haploid individual for X linked traits. Therefore they can only have 1 allele at a time.

Allelic and Genotypic frequencies

The allelic and genotypic frequencies remain the same in case of males.

Thus, p(.-) = p(.)

Where,

$$p(.-) = \frac{n_{.-}}{N_{.-}}$$

Where, $N_{.-} = \sum n_{.-}$ is the sum total of all the genotypes in the genotype pool.

$n_{.-}$ is the number of individuals with a particular genotype in the genotype pool.

and,

$$p(.) = \frac{n_.}{N_.}$$

Where,

$N_. = \sum n_.$ is the sum total of all the alleles in the gene pool.

$n_.$ is the number of particular alleles in the gene pool.

Therefore,

$$p(A-) = p(A) = \frac{n_{A-}}{N_{.-}} \Rightarrow \frac{400}{500} = 0.8$$

$$p(B-) = p(B) = \frac{n_{B-}}{N_{.-}} \Rightarrow \frac{100}{500} = 0.2$$

6.2.2. k alleles

The genetic structure of a population for a X linked trait with "k" alleles can be deduced separately for the female and male population. The genetic structure for the female population X linked trait with "k" alleles will be like that of a population with a single locus autosomal trait with "k" alleles already discussed earlier.

The tabulation for deducing gene and genotypic frequencies for a single allosomal locus for a X linked trait with "k" alleles (A1,A2, ...,Ak) in male population will be as follows.

S.No.	A1	A2	.	Ak	Genotypes	n
1	1	0	0	0	A1-	n_{A1-}
2	0	1	0	0	A2-	n_{A2-}
.	.	.	.	.	.	.
.	.	.	.	.	.	.
k	0	0	0	1	Ak-	n_{Ak-}
					Sum	$N_{.-}$

Here, k alleles A1, A2 ..... Ak can be taken as k dummy variables that can take any dummy value between 0 and 1 such that in each row A1 + A2 + ...+ Ak will always be equal to one (1). Since male individuals will act as a haploid individual for X linked traits. Therefore they can only have 1 allele at a time.

Allelic and Genotypic frequencies

The allelic and genotypic frequencies remain the same in case of males.

Thus, p(.-) = p(.)

Where,

$$p(.-) = p(.) = \frac{n_{.-}}{N_{.-}}$$

Where, $N_{.-} = \sum n_{.-}$ is the sum total of all the genotypes in the genotype pool.

$n_{.-}$ is the number of individuals with a particular genotype in the genotype pool.

Therefore,

$$p(A1-) = p(A1) = \frac{n_{A1-}}{N_{.-}}$$

$$p(A2-) = p(A2) = \frac{n_{A2-}}{N_{.-}}$$

.

$$p(Ak-) = p(Ak) = \frac{n_{Ak-}}{N_{.-}}$$

6.3. Multiple autosomal loci with multiple alleles

Now we will deduce the genetic structure of population for a trait affected by different genes at multiple autosomal loci and bears multiple alleles.

Illustration 1: Allelic dominance and the hierarchy of individual loci for coat colour in dogs and cats[44]

[44] Korec, E., Hančl, M., Bydžovská, M. et al. Inheritance of coat colour in the cane Corso Italiano dog. BMC Genet 20, 24 (2019). https://doi.org/10.1186/s12863-019-0731-2

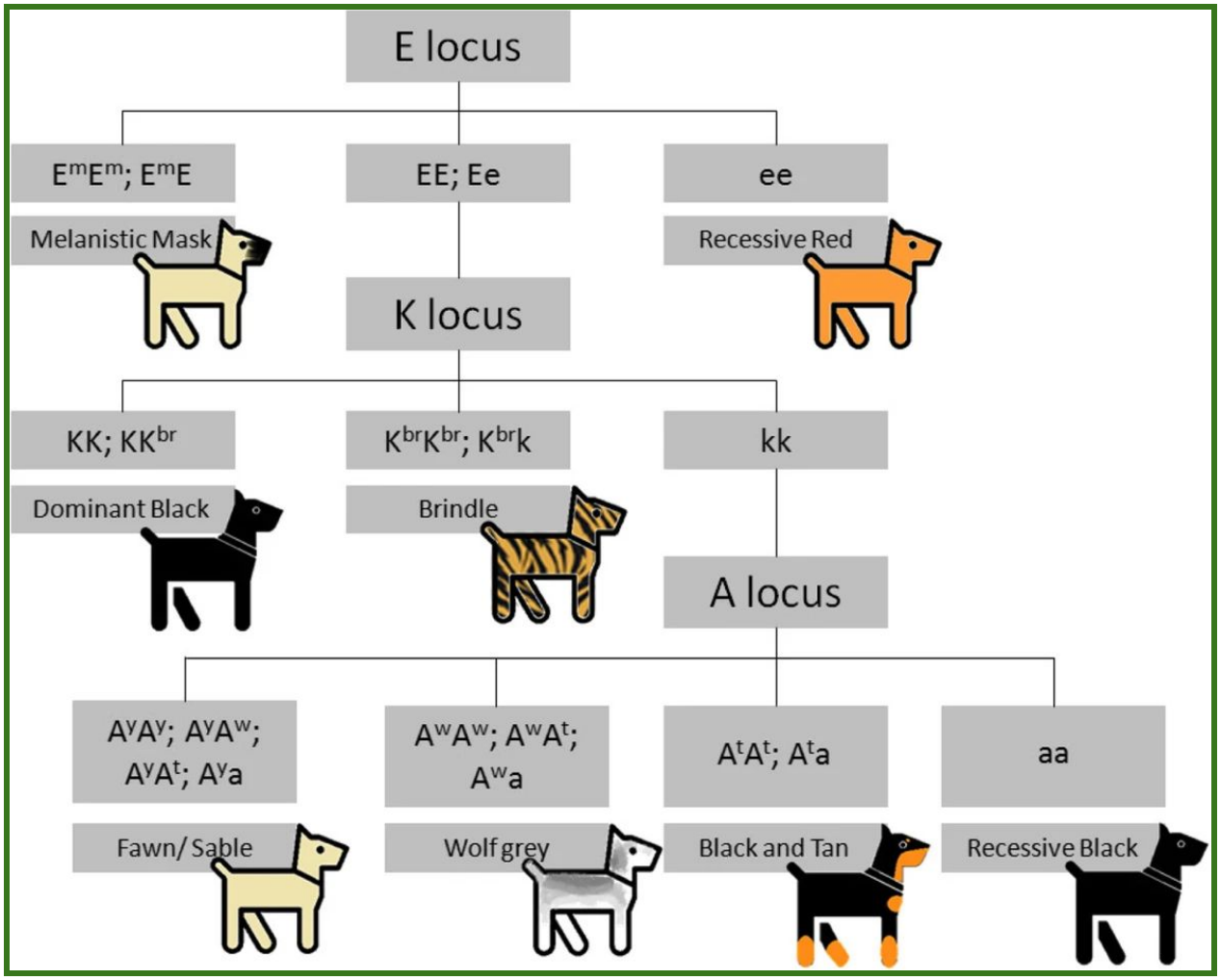

Fig 9. Inheritance of the basic coat colour loci in dogs and cats[45]

Locus E affects the distribution of eumelanin and pheomelanin. In dogs, there are three alleles with specific dominance ($E^m > E > e$). E^m is responsible for melanistic mask and e is responsible for recessive red. The K locus is hypostatic for the E locus, and there are also three alleles with a specific hierarchy ($K > K^{br} > k$). The K allele is responsible for dominant black and K^{br} is responsible for brindle colour. Locus A has four alleles, and the dominance of these alleles is $A^y > A^w > A^t > a$. Allele A^y is responsible for fawn or sable colour. Allele A^w represents wild colouration, which is ancestral. Allele A^t represents black and tan or saddle and tan colouration, and allele a is responsible for recessive black. Alleles E and k are wild-type, and coat colour is under the control of hypostatic

[45] Kaelin CB, Barsh GS. Genetics of pigmentation in dogs and cats. Annu Rev Anim Biosci. 2013;1:125–56.

loci.

Now we know that if there are k alleles, then number of possible genotypes in the population will be

$$= \frac{k(k+1)}{2}$$

ie. We have k1, k2 =3 for locus E and K, therefore, number of genotypes = $\frac{3(3+1)}{2} = 6$ each respectively

We have k3 =4 for locus A, therefore, number of genotypes = $\frac{4(4+1)}{2} = 10$

Therefore total number of genotypes considering three loci at a time = $6 \times 6 \times 10 = 360$

Generalized formula for calculating the total number of genotypes for " ψ " loci each having k1, k2, ... k ω alleles considering they assort independently will be

$$G = \frac{k1 \times k2 \times k3 \times \ldots\ldots\ldots \times k\omega \times (k1+1) \times (k2+1) \times (k3+1) \times \ldots\ldots\ldots \times (k\omega + 1)}{2^{\psi}}$$

For above illustration

$$= \frac{3 \times 3 \times 4 \times (3+1) \times (3+1) \times (4+1)}{2^3}$$

$$= \frac{3 \times 3 \times 4 \times 4 \times 4 \times 5}{2 \times 2 \times 2} = 360$$

Now solving a hypothetical example based on the above illustration with 360 genotypes is beyond the scope of this lecture notes. However one can very well accomplish this using computer softwares like excel etc. However, we can deduce a generalized methodology to understand how gene and genotypic frequencies can be calculated for multiple loci with multiple alleles.

The tabulation for deducing gene and genotypic frequencies for " ψ " loci with k1 alleles (1A1, 1A2, ..., 1Ak1), k2 alleles (2A1, 2A2, ..., 2Ak2), ... k ω alleles (ψ A1,

ψ A2, ..., ψ Ak ω) will be as follows.

S.No	**1A1**	**1A.**	**1Ak1**	**.A.**	ψ A1	ψ A.	ψ A *k*ω	**Genotypes**	**n**	**n × 1A1**	**n × 1A.**	**n × 1Ak1**	**n × .A.**	**n × ψ** A1	**n × ψ A.**	**n × ψ A** *k*ω
1	2	0	0	.	2	0	0	1A11A1 .... ψ A1 ψ A1	$n_{1A11A1....\psi A1\psi A1}$	$2\times n_{1A11A1....\psi A1\psi A1}$	$0\times n_{1A11A1....\psi A1\psi A1}$	$0\times n_{1A11A1....\psi A1\psi A1}$	$.\times n_{1A11A1....\psi A1\psi A1}$	$2\times n_{1A11A1....\psi A1\psi A1}$	$0\times n_{1A11A1....\psi A1\psi A1}$	$0\times n_{1A11A1....\psi A1\psi A1}$
.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.		.	.	.	.	.	.	.	.	.	.	.
G	0	0	2	.	0	0	2	1Ak11Ak1.... ψ A *k*ω ψ A *k*ω	$n_{1Ak11Ak1....\psi A k\omega\ \psi A k\omega}$	$0\times n_{1Ak11Ak1....\psi A k\omega\ \psi A k\omega}$	$0\times n_{1Ak11Ak1....\psi A k\omega\ \psi A k\omega}$	$0\times n_{1Ak11Ak1....\psi A k\omega\ \psi A k\omega}$	$0\times n_{1Ak11Ak1....\psi A k\omega\ \psi A k\omega}$	$0\times n_{1Ak11Ak1....\psi A k\omega\ \psi A k\omega}$	$0\times n_{1Ak11Ak1....\psi A k\omega\ \psi A k\omega}$	$0\times n_{1Ak11Ak1....\psi A k\omega\ \psi A k\omega}$
								Sum	$N_{1....G}$	n_{1A1}	$n_{1A.}$	n_{1Ak1}	$n_{.A.}$	$n_{\psi A1}$	$n_{\psi A.}$	$n_{\psi A\omega}$

Here, k1 alleles for first loci 1A1, 1A2 ..... 1Ak1 can be taken as k1 dummy variables that can take any dummy value among 0, 1 and 2 such that in each row 1A1 + 1A2 + ...+ 1Ak1 will always be equal to two (2). Since a diploid individual at a time can only have 2 alleles out of all possible alleles at a single loci. Similarly, ψ^{th} loci with ω alleles ψ A1, ψ A2, ..., ψ Ak ω can be taken as k ω dummy variables that can take any dummy value among 0, 1 and 2 such that in each row ψ A1 + ψ A2 + ...+ ψ Ak ω will always be equal to two (2).

Genotypic frequencies

Here, sum total of all possible genotypes in genotype pool will be $N_{..}$ and,

$$N_{.....\psi} = n_{1A11A1....\psi A1\psi A1} + ...+ n_{1Ak11Ak1....\psi A k\omega\ \psi A k\omega}$$

Now, genotypic frequency f(..) is the relative frequency of that genotype in the genotype pool and can alternatively be termed as probability of occurrence of a particular genotype in the genotype pool.

Formula for calculating genotypic frequencies in a diploid mendelian population for a ψ loci with k1, k2, ... k ω alleles can be given by

$$p(1A11A1....\psi A1\psi A1) = \frac{n_{1A11A1....\psi A1\psi A1}}{N_{1....G}}$$

.

.

$$p(1Ak11Ak1....\psi A k\omega\ \psi A k\omega) = \frac{n_{1Ak11Ak1....\psi Ak\omega\psi Ak\omega}}{N_{1....G}}$$

Where, $N_{1....G}$ = is the sum total of all the genotypes in the genotype pool.

$$p(1A11A1....\psi A1\psi A1) + ... + p(1Ak11Ak1....\psi A k\omega\ \psi A k\omega) = 1$$

Allelic frequencies

Now, from the table sum total of all the alleles in gene pool at single locus will be equivalent to $2N_{1....G}$ for diploid organisms. Also, allelic frequency f(.) is the relative frequency of that allele in the gene pool and can alternatively be termed as probability of occurrence of a particular allele in the gene pool.

Formula for calculating allelic frequencies in a diploid mendelian population for a particular alleles at a particular locus can be generalized as

$$p(.A.) = \frac{n_{.A.}}{N_{1.....G}} \quad \text{such that,}$$

$$\sum p(1A.) = p(1A1) + p(1A2) +.....+ p(1Ak1) = 1$$

$$\sum p(2A.) = p(2A1) + p(2A2) +.....+ p(2Ak2) = 1$$

.

.

$$\sum p(\psi A.) = p(\psi A1) + p(\psi A2) +.....+ p(\psi Ak\omega) = 1$$

7. Summary

In this lecture notes we understood basic concepts, like biological organization, population and its types, Mendelian population. Methodologies to calculate gene and genotypic frequencies in a population have been explained for autosomal and allosomal locus. Tabulation and methodology using dummy variables to calculate gene and genotypic frequencies for single locus two allele systems to that of multiple locus and multiple alleles have been undertaken.

****

ABOUT THE AUTHOR

Dr. Kuldeep Kumar Tyagi had completed his B.V.Sc & A.H. in the year 2006 from Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab India. He got admission in a master program in the subject of Animal Genetics and Breeding at Indian Veterinary Research Institute, Bareilly, Uttar Pradesh, India after securing 6th rank in All India ICAR-JRF examination. He had completed his Masters in the year 2008 and carried out research on competent fibroblast cells used in somatic cell nuclear transfer. He qualified CSIR Net in his first attempt during the final semester of masters program itself. He got selected as Assistant Professor in the year 2009 at College of Veterinary Science & A.H. at Navsari Agricultural University, Navsari, Gujarat, India. He enriched his practical knowledge and expertise in the subject of Animal Breeding while disbursing his duties as Scheme Incharge at Livestock Research Station of the same university for 9 years. During the same tenure he also accumulated practical expertise on various aspects of field level breeding programs while heading "All India Coordinated Research Project on Goat Improvement - Surti Field Unit" as Principal Investigator. He completed his Ph.D. in the year 2016 from the same university as an inservice candidate. He had worked on gene expression studies on mammary epithelial cells of buffaloes during his Ph.D. degree program. He had been selected as Associate Professor in the department of Animal Genetics & Breeding, College of Veterinary and Animal Science, Sardar Vallbhbhai, Patel University of Agriculture & Technology, Meerut, Uttar Pradesh, India in the year 2018. Since then he has been heading the same department as Officer-Incharge. He had handled 5 externally funded and 27 institutionally funded research projects. He had coguided two masters students. He has in his credit 61 research papers, 14 research recommendations, 5 lecture notes and 4 success stories. He is a member of 4 professional societies and attended 21 conferences/ symposiums/ workshops. He has remained on a panel of experts for framing question papers for various Universities and National level examination bodies. He is hosting a google site for online teaching https://sites.google.com/view/learnagb and can be reached at drtyagivet@gmail.com for initiating a conversation.

Published by:

Department of Animal Genetics & Breeding
College of Veterinary & Animal Sciences
Sardar Vallabhbhai Patel University of Agriculture and Technology
Meerut- 250 110, Uttar Pradesh, India

To cite this lecture notes:

Tyagi, K 2021, *Genetic Structure of Population*, lecture notes, Principles of Animal and Population Genetics AGB UNIT II, Sardar Vallabhbhai Patel University of Agriculture & Technology, Meerut, India, Delivered 19, 20 & 23 February 2021. Retrieved online from https://vepub.com