QG Models 4

Authors: Author 1, Author 2, ..., H. Jorjani
Last update: 2020 02 19

Some title

Some sentences.

Some author

In this chapter, the central dogma of QG, represented in the conceptual equation $P = G + E$, is redefined in form of the more operationally (statistically) suitable equation $\mathbf{y} = \mathbf{X}\beta + \mathbf{Zu} + \epsilon$. Afterwards, assumptions of the classical QG theory are briefly described. Departures from these assumptions can, in most cases, be handled as minor anomalies within the framework of the classical QG theory by treating them as nuisance parameters. However, some of the departures have been handled by making them a parameter of interest, and include them specifically in the model.

4.1 Central dogma of QG

The central dogma of the classical quantitative genetics, attributed to the founders of quantitative genetics, Fisher (1918), Wright (1921), and Haldane (1924) (see also Haldane (1932)[1]), is nowadays commonly written as:

$$P = G + E \tag{4.1}$$

where,

P = the phenotypic value,
G = the genotypic value, and
E = the environmental value.

Founders of the quantitative genetics did not necessarily use the symbols P, G, and E, or the short definitions mentioned above. For example, Fisher (1918) used the words *"feature"* and *"measurement"* instead of the phenotypic value. Further, Fisher did not use the words genotypic value, instead he used the words *"genetic value"* and *"genic value"* (the definitions of which will be presented shortly). Furthermore, he defined environment as *"arbitrary external causes independent of heredity"*. Wright (1921a) did not explicitly use the symbol P in any equation, however the words phenotype and genotype were used by him. With regard to the environment, Wright partitioned environment into a *"tangible"* part (with

1: **Haldane's contributions to QG theory:** Contributions of Haldane to QG theory is perhaps best described as *"diffused"*, in the sense that it is difficult to pinpoint anything that is directly quantitative genetics in his massive publications from (1924) and onward. However, many elements of his contributions to population genetics theory gradually found their ways to QG theory. His earlier contributions are well summarized in his book, entitled *"causes of evolution"* (1932). It is from this book that his many significant contribution to QG theory can be observed.

symbol E), and an *"intangible"* part (with the symbol D). In Haldane (1924) the word genotype was not used, and the word *"character"* was used instead of phenotype. So, it can easily be concluded the the symbols and what they stand for is just the result of a tacit agreement among the the people using them since 1920's. There is nothing sacrosanct about them. We will see later that we may have to change many symbols to bring more clarity into the discussion.

The first thing to note about the model in Equation 4.1 is that it is an additive model. It can be speculated, as Gianola* has done, that alternative specific models such as $P = G^E$, $P = E^G$, or $P = (G + E)^{GE}$ could have been suggested. The simple equation $P = G + E$ has some strong underlying assumptions (which will be discussed in the next section). The minimum that founders of QG theory could do was to start with an assumption-free equation such as $P = f(G, E)$, or $P \propto G, E$. However, in the light of the history of QG theory (see Section 3.2 and Section 3.3), the use of an additive model is understandable. In other words, an additive model is a natural consequence of invoking the central limit theorem and normal distribution [2] .

2: **Central Limit Theorem and normal distribution:** Central Limit Theorem and normal distribution will be discussed in more details in Section 5.1.

Another aspect of Equation 4.1 that is not clear is whether P, G, and E are referring to an individual or to a population. As long as $P = G + E$ is a conceptual definition, it does not matter so much if, *e.g.* P refers to an individual or a population. However, if the purpose of the model is to describe the state of a population, perhaps it it is better to write Equation 4.1 as:

$$P_i = G_i + E_i \tag{4.2}$$

Equation 4.2 is still a conceptual equation, and if we want it to be operationally useful, *i.e.* to be a statistically usable model, then Equation 4.2 needs to be modified by adding a residual (or an error) term, using the symbol ϵ, as in:

$$P_i = G_i + E_i + \epsilon_i \tag{4.3}$$

We can see that Equation 4.3 conforms better to the specifications suggested by Wright (1921a) who separated the environment into a *"tangible"* part, here symbolized by E, and an *"intangible"* part, here symbolized by ϵ. In the modern view, the term ϵ is not restricted only to the *"intangible"* part of the environment, but it is also pertinent to all uncertainties in the model. In other words, ϵ is a reflection of the factors that we have neglected to include in the model, and our ignorance of all other factors that affect the main variables of interest, *i.e.* P, G, and E.

A tacit assumption in Equation 4.1 to Equation 4.3 is that all elements are expressed as deviations from their mean value. If we express each element as the phenotype is measured, then we need to add a population mean (μ) to the equation:

$$P_i = \mu + G_i + E_i + \epsilon_i \tag{4.4}$$

* Kindly provided by D. Gianola (2015) from a presentation entitled: *A brief history of statistical methods in animal breeding*. See also the presentation by Gianola at the HSTalks: https://hstalks.com/t/3440/a-brief-history-of-statistical-developments-in-ani/.

We can already see that the equation representing the central dogma of QG theory is growing in complexity from Equation 4.1 to Equation 4.4, and in future discussions the notation may become too complex to render clarity. Using matrix notation may facilitate further development of the equations. In matrix notation, Equation 4.4 can be written as:

$$\mathbf{P} = \mu + \mathbf{G} + \mathbf{E} + \epsilon \tag{4.5}$$

Elements of Equation 4.5 deserve separate discussions. Let's first direct our attention to E_i, *i.e.* the environments that we are explicitly modeling. An underlying assumption of Equation 4.1 to Equation 4.5, is that all individuals, or at least all individuals with the same genotype for the trait under study, are randomly distributed across all environments. But this is a very strong assumption. Conceptually, one can think of a population consisting of an infinite number of individuals, for any specific genotype, that are equally distributed in every conceivable type of environment. But this is unrealistic. In reality, and as an example, dairy cows are unique, they do not come in an infinite number of copies (clones), and any cow is confined to only one environment (herd). Therefore, we need to specify which individual has been in which environment. This can be done by means of an incidence matrix (here symbolized by X)[3] .

As interesting as the explicitly modeled environmental factors may be, they are probably not the main focus of a QG analysis. The same is true of the population mean. So, we may treat them as nuisance parameters[4] , and include them in a column vector called β, where $\beta = [\mu \quad E]'$, and the also use the incidence matrix X, to relate each element of P to different elements of β. Consequently the equation becomes:

$$\mathbf{P} = \mathbf{X}\beta + \mathbf{G} + \epsilon \tag{4.6}$$

There are also some assumptions related to the element G (the genotypic value) in the above equations. From these equations it is not clear if the values of P for different individual are independent of each other or somehow they are related to each other through G. To allow for a proper specification of this relationship, it is better to use an incidence matrix (here symbolized as Z) that relates different values of P to G. At the same time, and in order to make the distinction clear, we may use a new symbol (u) instead of G. Consequently, Equation 4.6 changes to:

$$\mathbf{P} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \epsilon \tag{4.7}$$

Finally, both in statistics and in animal/plant breeding literature, the (phenotypic) observations are usually shown by another letter[5] , namely y. Therefore, in order to create harmony between the notations from different disciplines, it is better to re-write Equation 4.6 as:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \epsilon \tag{4.8}$$

Comparing Equations 4.1 and 4.9, we can see that we have moved from a conceptual definition of QG theory's central dogma to an operational definition that easily lends itself to statistical analyses of data relevant

3: **Incidence matrix:** In statistical genetics, the incidence matrix, also known as the design matrix, is a matrix consisting only of 0 and 1, with one row for each observation, and one column for each level of the modeled effects. For example, if there are 4 cow observations from 2 herds, the incidence matrix X will have 4 rows for 4 cows, and 3 columns for the population mean and the 2 herds.

4: **Nuisance parameter:** In statistics, a nuisance parameter is a parameter that is not of primary interest.

5: **Symbols have no inherent meaning:** In the evolutionary quantitative genetics, instead of P or y, usually the symbol z is used (see *e.g.* Lynch and Walsh (1998) and Walsh and Lynch (2018)). As mentioned before, there is nothing sacrosanct about symbols.

to QG studies. Additionally, Equation 4.9 removes some opacity of the underlying assumptions, as well as providing the means to specify different assumptions and models in a convenient way. This will also make it easier to refer *estimation* and *prediction* issues to the literature specialized in such areas (*e.g.* Mrode (2013), Schaeffer (2019), and Sorensen and Gianola (2002)).

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \epsilon \tag{4.9}$$

4.2 Classical Quantitative Genetic Model

Predecessors of Fisher, such as Yule (1902, 1906) and Pearson (1904a), in their efforts to reconcile the continuous variation with the Mendelian genetics made some implicit or explicit assumptions, which after modifications by Fisher (1918), can be summarized as follows:

1. There are n bi-allelic loci, each with a small effect on the variance of the trait under under consideration;
2. Alleles within a locus may may have any frequency, show non-additive (*i.e.* dominance/recessive) relationship, and non-additivity may be of any degree;
3. Alleles across loci may show non-additive (*i.e.* epistatic) relationship, and non-additivity may be of any degree;
4. Non-additivity and environment reduce the correlation between relatives.

4.2.1 Central Limit Theorem

A summary of this chapter.

Bibliography

Citations in alphabetical order.

Alberts, B., A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter (2015). *Molecular biology of the cell*. Sixth edition. New York, NY: Garland Science, Taylor and Francis Group. 1465 pp. (cited on page 5).

Bagheri, H. C. and G. P. Wagner (2004). Evolution of Dominance in Metabolic Pathways. *Genetics* **168**: 1713–1735. DOI: 10.1534/genetics.104.028696 (cited on page 11).

Bateson, W. (1901). Problems of heredity as a subject for horticultural investigations. *Journal of the Royal Horticultural Society* **25**: 54–61. DOI: 10.1017/CBO9780511693946 (cited on page 4).

Bolormaa, S., J. E. Pryce, B. J. Hayes, and M. E. Goddard (2010). Multivariate analysis of a genome-wide association study in dairy cattle. *Journal of Dairy Science* **93**: 3818–3833. DOI: 10.3168/jds.2009-2980 (cited on page 13).

Bolormaa, S., J. E. Pryce, A. Reverter, Y. Zhang, W. Barendse, K. Kemper, B. Tier, K. Savin, B. J. Hayes, and M. E. Goddard (2014). A Multi-Trait, Meta-analysis for Detecting Pleiotropic Polymorphisms for Stature, Fatness and Reproduction in Beef Cattle. *PLoS Genetics* **10**: ed. by J. Flint, e1004198. DOI: 10.1371/journal.pgen.1004198 (cited on page 13).

Bromham, L. (2016). What is a gene for? *Biology & Philosophy* **31**: 103–123. DOI: 10.1007/s10539-014-9472-9 (cited on page 12).

Bulmer, M. G. (2003). *Francis Galton: pioneer of heredity and biometry*. Baltimore: Johns Hopkins University Press. 357 pp. (cited on page 17).

Bünger, L., U. Renne, G. Dietl, and S. Kuhla (1998). Long-term selection for protein amount over 70 generations in mice. *Genetical Research* **72**: 93–109. DOI: 10.1017/S0016672398003401 (cited on pages 13, 14).

Casella, G. and R. L. Berger (2002). *Statistical inference*. 2nd Ed. Duxbury. 686 pp. (cited on page 6).

Castle, W. E. (1905). The Mutation Theory of Organic Evolution, from the Standpoint of Animal Breeding. *Science* **21**: 521–525. DOI: 10.1126/science.21.536.521 (cited on page 19).

Chang, C.-W., W.-C. Cheng, C.-R. Chen, W.-. Shu, M.-L. Tsai, C.-L. Huang, and I. C. Hsu (2011). Identification of Human Housekeeping Genes and Tissue-Selective Genes by Microarray Meta-Analysis. *PLOS ONE* **6**: e22859. DOI: 10.1371/journal.pone.0022859 (cited on page 11).

Cohen, I. B. (1985). *Revolution in science*. 1st Ed. Harvard University Press. 746 pp. (cited on page viii).

Collas, P., T. M. Liyakat Ali, A. Brunet, and T. Germier (2019). Finding Friends in the Crowd: Three-Dimensional Cliques of Topological Genomic Domains. *Frontiers in Genetics* **10**: 602. DOI: 10.3389/fgene.2019.00602 (cited on page 5).

Crow, J. F. and M. Kimura (2009). *An introduction to population genetics theory*. OCLC: 1027901624. Jodhpur; New Jersey: Scientific Publisher (India) ; The Blackburn Press (cited on pages 7, 13, 16).

Darwin, C. R. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London. 502 pp. (cited on pages 15, 16, 18, 19).

Dean, A. M., D. E. Dykhuizen, and D. L. Hartl (1986). Fitness as a function of ?-galactosidase activity in *Escherichia coli*. *Genetical Research* **48**: 1–8. DOI: 10.1017/S0016672300024587 (cited on page 11).

Dudley, J. W. (2007). From Means to QTL: The Illinois Long-Term Selection Experiment as a Case Study in Quantitative Genetics. *Crop Science* **47**: (Supplement_3), S–20. DOI: 10.2135/cropsci2007.04.0003IPBS (cited on page 13).

Dudley, J. W. and R. J. Lambert (2010). '100 Generations of Selection for Oil and Protein in Corn'. *Plant Breeding Reviews*. Oxford, UK: John Wiley & Sons, Inc., pp. 79–110. DOI: 10.1002/9780470650240.ch5 (cited on page 14).

Dunnington, E. A., C. F. Honaker, M. L. McGilliard, and P. B. Siegel (2013). Phenotypic responses of chickens to long-term, bidirectional selection for juvenile body weight: A Historical perspective. *Poultry Science* **92**: 1724–1734. DOI: 10.3382/ps.2013-03069 (cited on page 13).

Dykhuizen, D. E., A. M. Dean, and D. L. Hartl (1987). Metabolic Flux and Fitness. *Genetics* **115**: 25–31 (cited on page 11).

East, E. M. (1910). A Mendelian Interpretation of Variation that is Apparently Continuous. *The American Naturalist* **44**: 65–82 (cited on page 19).
Eisenberg, E. and E. Y. Levanon (2013). Human housekeeping genes, revisited. *Trends in Genetics* **29**: 569–574. DOI: 10.1016/j.tig.2013.05.010 (cited on page 11).
Eldredge, N. and S. J. Gould (1972). 'Punctuated equilibria: an alternative to phyletic gradualism'. *Models in paleobiology*. Cooper and Co., San Francisco, pp. 82–115 (cited on page 19).
Ewens, W. J. (2004). *Mathematical population genetics. 1: Theoretical introduction*. 2. ed. Interdisciplinary applied mathematics Mathematical biology. New York, NY: Springer. 417 pp. (cited on page 7).
Falconer, D. S. (1960). *Introduction to Quantitative Genetics*. 1st edition (cited on page v).
Falconer, D. S. and T. F. C. Mackay (1996). *Introduction to Quantitative Genetics*. 4th ed. Longman Group Ltd. 480 pp. (cited on pages v, vi, viii).
Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* **525**: 399–433 (cited on pages 4, 10, 20, 23, 26).
Galton, F. (1865). Hereditary talent and character. *Macmillan's Magazine* **12**: 157–166, 318–327 (cited on page 18).
Galton, F. (1877). Typical laws of heredity, 492–533 (cited on page 18).
Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland* **15**: 246–263 (cited on pages 4, 19).
Galton, F. (1889). *Natural inheritance*. Macmillan, London (cited on page 19).
Geiger, T., A. Wehner, C. Schaab, J. Cox, and M. Mann (2012). Comparative Proteomic Analysis of Eleven Common Cell Lines Reveals Ubiquitous but Varying Expression of Most Proteins. *Molecular & Cellular Proteomics : MCP* **11**: DOI: 10.1074/mcp.M111.014050 (cited on page 11).
Gould, S. J. (2002). *The structure of evolutionary theory*. Cambridge, Mass: Belknap Press of Harvard University Press. 1433 pp. (cited on page 19).
Griffiths, A. J., S. R. Wessler, J. Carroll, and J. Doebley (2015). *Introduction to genetic analysis*. 11th Ed. W. H. Freeman & Company (cited on pages 5, 15).
Haldane, J. B. S. (1924). A mathematical theory of natural and artificial selection. Part I. (cited on pages 23, 24).
Haldane, J. B. S. (1932). *The causes of evolution*. Longmans, Green and Co. 235 pp. (cited on page 23).
Hamilton, M. B. (2009). *Population genetics*. OCLC: ocn259716125. Chichester, UK ; Hoboken, NJ: Wiley-Blackwell. 407 pp. (cited on page 7).
Hartl, D. L. and A. G. Clark (2007). *Principles of population genetics*. 4th Ed. Sinauer Associates (cited on page 7).
Hartl, D. L., D. E. Dykhuizen, and A. M. Dean (1985). Limits of Adaptation: The Evolution of Selective Neutrality. *Genetics* **111**: 655–674 (cited on page 11).
Hill, W. G. (2005). A Century of Corn Selection. *Science* **307**: 683–685 (cited on page 14).
Hill, W. G. (1984a). *Quantitative Genetics: Explanation and analysis of continuous variation*. Van Nostrand Reinhold. 376 pp. (cited on page 21).
Hill, W. G. (1984b). *Quantitative genetics: Selection*. Van Nostrand Reinhold. 426 pp. (cited on page 21).
Hofmeyr, J.-H. S. and A. Cornish-Bowden (1991). Quantitative assessment of regulation in metabolic systems. *European Journal of Biochemistry* **200**: 223–236. DOI: 10.1111/j.1432-1033.1991.tb21071.x (cited on page 11).
Hogg, R. V., J. W. McKean, and A. T. Craig (2019). *Introduction to mathematical statistics*. Eighth edition. Boston: Pearson. 746 pp. (cited on page 6).
Holt, M., T. Meuwissen, and O. Vangen (2005). Long-term responses, changes in genetic variances and inbreeding depression from 122 generations of selection on increased litter size in mice. *Journal of Animal Breeding and Genetics* **122**: 199–209. DOI: 10.1111/j.1439-0388.2005.00526.x (cited on page 13).
Houle, D., D. K. Hoffmaster, S. Assimacopoulos, and B. Charlesworth (1992). The genomic mutation rate for fitness in Drosophila. *Nature* **359**: 58–60. DOI: 10.1038/359058a0 (cited on page 13).
Hubby, J. L. and R. C. Lewontin (1966). A Molecular Approach to the Study of Genic Heterozygosity in Natural Populations. I. the Number of Alleles at Different Loci in Drosophila Pseudoobscura. *Genetics* **54**: 577–594 (cited on page 11).
Hull, D. L. (1988). *Science as a process: an evolutionary account of the social and conceptual development of science*. Paperback ed., [Nachdr.] Science and its conceptual foundations. OCLC: 837696311. Chicago: University of Chicago Press. 586 pp. (cited on page 15).

Jensen, J. D., B. A. Payseur, W. Stephan, C. F. Aquadro, M. Lynch, D. Charlesworth, and B. Charlesworth (2019). The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018: COMMENTARY. *Evolution* **73**: 111–114. DOI: 10.1111/evo.13650 (cited on page 18).

Jiang, J., L. Ma, D. Prakapenka, P. M. VanRaden, J. B. Cole, and Y. Da (2019). A Large-Scale Genome-Wide Association Study in U.S. Holstein Cattle. *Frontiers in Genetics* **10**: 412. DOI: 10.3389/fgene.2019.00412 (cited on page 13).

Johansson, A. M., M. E. Pettersson, P. B. Siegel, and Ö. Carlborg (2010). Genome-Wide Effects of Long-Term Divergent Selection. *PLoS Genetics* **6**: ed. by B. Walsh, e1001188. DOI: 10.1371/journal.pgen.1001188 (cited on page 14).

Kacser, H. (1989). 'Quantitative variation and the control analysis of enzyme systems.' *Hill W. G., and T.F.C. Mackay (1989) Evolution and Animal Breeding: Reviews on Molecular and Quantitative Approaches in Honour of Alan Robertson*. C.A.B. International, Wallingford, UK, pp. 219–226 (cited on page 11).

Kacser, H. and J. A. Burns (1979). Molecular Democracy: Who Shares the Controls? *Biochemical Society Transactions* **7**: 1149–1160. DOI: 10.1042/bst0071149 (cited on page 11).

Kacser, H. and J. A. Burns (1981). The Molecular Basis of Dominance. *Genetics* **97**: 639–666 (cited on pages 5, 11).

Keel, B. N., D. J. Nonneman, A. K. Lindholm-Perry, W. T. Oliver, and G. A. Rohrer (2019). A Survey of Copy Number Variation in the Porcine Genome Detected From Whole-Genome Sequence. *Frontiers in Genetics* **10**: 737. DOI: 10.3389/fgene.2019.00737 (cited on page 5).

Keightley, P. D. and W. G. Hill (1992). Quantitative Genetic Variation in Body Size of Mice From New Mutations. *Genetics* **131**: 693–700 (cited on page 13).

Kempthorne, O. (1976). 'Status of quantitative genetics'. *Proceedings of the International Conference on Quantitative Genetics*. Iowa State University Press, pp. 719–760 (cited on page 4).

Kern, A. D. and M. W. Hahn (2018). The Neutral Theory in Light of Natural Selection. *Molecular Biology and Evolution* **35**: ed. by S. Kumar, 1366–1371. DOI: 10.1093/molbev/msy092 (cited on page 18).

Kimura, M. (1968). Evolutionary Rate at the Molecular Level. *Nature* **217**: 624–626. DOI: 10.1038/217624a0 (cited on pages 11, 17).

Kuhn, T. S. (1996). *The structure of scientific revolutions*. 3rd Ed. Chicago, IL: University of Chicago Press. 212 pp. (cited on page viii).

Larsen, R. J. and M. L. Marx (2017). *An Introduction to Mathematical Statistics and Its Applications* (cited on page 6).

Laurie, C. C., S. D. Chasalow, J. R. LeDeaux, R. McCarroll, D. Bush, B. Hauge, C. Lai, D. Clark, T. R. Rocheford, and J. W. Dudley (2004). The Genetic Architecture of Response to Long-Term Artificial Selection for Oil Concentration in the Maize Kernel. *Genetics* **168**: 2141–2155. DOI: 10.1534/genetics.104.029686 (cited on page 14).

Lehmann, E. L. and G. Casella (1998). *Theory of point estimation*. 2nd Ed. Springer texts in statistics. New York: Springer. 589 pp. (cited on page 6).

Lehninger, A. L., D. L. Nelson, and M. M. Cox (2013). *Lehninger principles of biochemistry*. 6th ed. OCLC: ocn820352899. New York: W.H. Freeman. 1119 pp. (cited on page 5).

Lewontin, R. C. and J. L. Hubby (1966). A Molecular Approach to the Study of Genic Heterozygosity in Natural Populations. Ii. Amount of Variation and Degree of Heterozygosity in Natural Populations of Drosophila Pseudoobscura. *Genetics* **54**: 595–609 (cited on page 11).

Liu, Y., X. Chang, J. Glessner, H. Qu, L. Tian, D. Li, K. Nguyen, P. M. A. Sleiman, and H. Hakonarson (2019). Association of Rare Recurrent Copy Number Variants With Congenital Heart Defects Based on Next-Generation Sequencing Data From Family Trios. *Frontiers in Genetics* **10**: 819. DOI: 10.3389/fgene.2019.00819 (cited on page 5).

Liu, Y. and X. Li (2014). Has Darwin's Pangenesis Been Rediscovered? *BioScience* **64**: 1037–1041. DOI: 10.1093/biosci/biu151 (cited on page 17).

Lynch, M. and B. Walsh (1998). *Genetics and analysis of quantitative traits*. Sunderland, Mass: Sinauer. 980 pp. (cited on pages vi, 25).

Mackay, T. F. C., J. D. Fry, R. F. Lyman, and S. V. Nuzhdin (1994). Polygenic mutation in Drosophila melanogaster: estimates from response to selection of inbred strains. *Genetics* **136**: 937–951 (cited on page 13).

Magdeldin, S., S. Enany, Y. Yoshida, B. Xu, Y. Zhang, Z. Zureena, I. Lokamani, E. Yaoita, and T. Yamamoto (2014). Basics and recent advances of two dimensional- polyacrylamide gel electrophoresis. *Clinical proteomics* **11**: 16. DOI: 10.1186/1559-0275-11-16 (cited on page 11).

Maindonald, J. and W. J. Braun (2010). *Data Analysis and Graphics Using R - an Example-Based Approach*. 3rd Ed. Cambridge University Press (cited on page 6).

Malthus, T. (1798). *An Essay on the Principle of Population*. J. Johnson. 125 pp. (cited on pages 16, 17).

Marouli, E. *et al.* (2017). Rare and low-frequency coding variants alter human adult height. *Nature* **542**: 186–190. DOI: 10.1038/nature21039 (cited on page 12).

Mayr, E. (1977). Darwin and Natural Selection: How Darwin may have discovered his highly unconventional theory. *American Scientist* **65**: 321–327 (cited on pages 15, 16).

Mayr, E. (1982). *The growth of biological thought: diversity, evolution, and inheritance*. Cambridge, Mass.: Harvard Univ. Pr. 974 pp. (cited on page 21).

Mayr, E. (1988). *Toward a new philosophy of biology: Observations of an evolutionist*. Harvard University Press. 564 pp. (cited on page 15).

Meaburn, K. J. and T. Misteli (2019). Assessment of the Utility of Gene Positioning Biomarkers in the Stratification of Prostate Cancers. *Frontiers in Genetics* **10**: 1029. DOI: 10.3389/fgene.2019.01029 (cited on page 5).

Mendel, G. (1866). Versuche über Plflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn*, 3–47 (cited on pages 4, 19).

Moose, S. P., J. W. Dudley, and T. R. Rocheford (2004). Maize selection passes the century mark: a unique resource for 21st century genomics. *Trends in Plant Science* **9**: 358–364. DOI: 10.1016/j.tplants.2004.05.005 (cited on page 14).

Mrode, R. A. (2013). *Linear models for the prediction of animal breeding values*. 3rd ed. Boston, MA: CABI (cited on pages 6, 26).

Nielsen, R. and M. Slatkin (2013). *An Introduction to Population Genetics: Theory and Applications*. Oxford, New York: Oxford University Press. 287 pp. (cited on page 7).

O'Farrell, P. H. (1975). High Resolution Two-Dimensional Electrophoresis of Proteins. *The Journal of biological chemistry* **250**: 4007–4021 (cited on page 11).

Pearson, K. (1904a). Report on Certain Enteric Fever Inoculation Statistics. *BMJ* **2**: 1243–1246. DOI: 10.1136/bmj.2.2288.1243 (cited on pages 19, 26).

Pearson, K. (1904b). III . Mathematical contributions to the theory of evolution. XII. On a generalised Theory of alternative Inheritance, with special reference to Mendel's laws. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **203**: 53–86. DOI: 10.1098/rsta.1904.0015 (cited on page 19).

Perason, K. and A. Lee (1903). On the Laws of Inheritance in Man: I. Inheritance of Physical Characters. *Biometrika* **2**: 357–462 (cited on page 19).

Pierce, B. A. (2012). *Genetics: A conceptual approach*. 4th ed. New York: W.H. Freeman. 1 p. (cited on page 5).

Portin, P. and A. Wilkins (2017). The Evolving Definition of the Term "Gene". *Genetics* **205**: 1353–1364. DOI: 10.1534/genetics.116.196956 (cited on page 12).

Price, G. R. (1970). Selection and covariance. *Nature* **227**: 520–521 (cited on page 7).

Robertson, A. (1966). A mathematical model of the culling process in dairy cattle. *Animal Science* **8**: 95–108. DOI: 10.1017/S0003356100037752 (cited on page 7).

Rogers, S. O. (2017). *Integrated Molecular Evolution* (cited on page 12).

Rowley, M. J. and V. G. Corces (2018). Organizational principles of 3D genome architecture. *Nature Reviews Genetics* **19**: 789–800. DOI: 10.1038/s41576-018-0060-8 (cited on page 5).

Ruse, M. (2014). 'Gould, Stephen Jay'. *eLS. John Wiley & Sons, Ltd: Chichester.* American Cancer Society, pp. 1–10. DOI: 10.1002/9780470015902.a0025067 (cited on page 19).

Sanders, M. F. and J. L. Bowman (2015). *Genetic Analysis - An Integrated Approach*. OCLC: 915123587 (cited on pages 5, 15).

Santiago, E., J. Albornoz, A. Dominguez, M. A. Torot, and C. Lopez-Fanjul (1992). The Distribution of Spontaneous Mutationson Quantitative Traits and Fitness in Drosophila melanogaster. *Genetics* **132**: 771–781 (cited on page 13).

Savage, J. E. *et al.* (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics* **50**: 912–919. DOI: 10.1038/s41588-018-0152-6 (cited on page 12).
Schaeffer, L. R. (2019). *Animal models*. L. R. Schaeffer. 381 pp. (cited on pages 6, 26).
Shannon, W. D., M. A. Watson, A. Perry, and K. Rich (2002). Mantel statistics to correlate gene expression levels from microarrays with clinical covariates. *Genetic Epidemiology* **23**: 87–96. DOI: 10.1002/gepi.1115 (cited on page 12).
Sorensen, D. and D. Gianola (2002). *Likelihood, Bayesian and MCMC methods in quantitative genetics*. Statistics for biology and health. New York: Springer-Verlag. 740 pp. (cited on pages 6, 26).
Strachan, T., J. Goodship, and P. F. Chinnery (2015). *Genetics and genomics in medicine*. New York: Garland Science/Taylor & Francis Group. 526 pp. (cited on page 5).
Terman, L. M. (1917). The Intelligence Quotient of Francis Galton in Childhood. *The American Journal of Psychology* **28**: 209–215 (cited on page 18).
Tymoczko, J. L., J. M. Berg, and L. Stryer (2015). *Biochemistry, a short course*. Third edition. New York: W.H. Freeman & Company, a Macmillan Education imprint. 900 pp. (cited on page 5).
Uhlén, M., L. Fagerberg, B. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, . Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. v. Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. v. Heijne, J. Nielsen, and F. Pontén (2015). Tissue-based map of the human proteome. *Science* **347**: DOI: 10.1126/science.1260419 (cited on page 11).
Visscher, P. M., N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, M. A. Brown, and J. Yang (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics* **101**: 5–22. DOI: 10.1016/j.ajhg.2017.06.005 (cited on page 12).
Wackerly, D. D., W. Mendenhall III, and R. L. Scheaffer (2008). *Mathematical Statistics*. 8th Ed. Brooks/Cole, Cengage Learning. 946 pp. (cited on page 6).
Walsh, B. and M. Lynch (2018). *Evolution and selection of quantitative traits*. New York, NY: Oxford University Press. 1459 pp. (cited on pages vi, 7, 25).
Watson, M. A., A. Perry, V. Budhjara, C. Hicks, W. D. Shannon, and K. M. Rich (2001). Gene Expression Profiling with Oligonucleotide Microarrays Distinguishes World Health Organization Grade of Oligodendrogliomas. *Cancer Research* **61**: 1825–1829 (cited on page 12).
Wilhelm, M., J. Schlegl, H. Hahne, A. M. Gholami, M. Lieberenz, M. M. Savitski, E. Ziegler, L. Butzmann, S. Gessulat, H. Marx, T. Mathieson, S. Lemeer, K. Schnatbaum, U. Reimer, H. Wenschuh, M. Mollenhauer, J. Slotta-Huspenina, J.-H. Boese, M. Bantscheff, A. Gerstmair, F. Faerber, and B. Kuster (2014). Mass-spectrometry-based draft of the human proteome. *Nature* **509**: 582–587. DOI: 10.1038/nature13319 (cited on page 11).
Wood, A. R. *et al.* (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* **46**: 1173–1186. DOI: 10.1038/ng.3097 (cited on page 12).
Wright, S. (1921a). Systems of mating: I-V. *Genetics* **6**: 111–178 (cited on pages 10, 23, 24).
Wright, S. (1921b). Systems of mating. I. The biometric relations between parent and offspring. *Genetics* **6**: 111–123 (cited on page 10).
Wright, S. (1921c). Systems of mating. II. The effects of inbreeding on the genetic composition of a population. *Genetics* **6**: 124–143 (cited on page 10).
Wright, S. (1921d). Systems of mating. III. Assortative mating based on somatic resemblance. *Genetics* **6**: 144–161 (cited on page 10).
Wright, S. (1921e). Systems of mating. IV. The effects of selection. *Genetics* **6**: 162–166 (cited on page 10).
Wright, S. (1921f). Systems of mating. V. General considerations. *Genetics* **6**: 167–178 (cited on page 10).
Yoo, B. H. (1980). Long-term selection for a quantitative character in large replicate populations of *Drosophila melanogaster* : 1. Response to selection. *Genetical Research* **35**: 1–17. DOI: 10.1017/S0016672300013896 (cited on page 13).
Yule, G. U. (1902). MENDEL'S LAWS AND THEIR PROBABLE RELATIONS TO INTRA-RACIAL HEREDITY. *New Phytologist* **1**: 222–238. DOI: 10.1111/j.1469-8137.1902.tb07336.x (cited on pages 19, 26).

Yule, G. U. (1906). 'On the theory of inheritance of quantitative compound characters on the basis of Mendels laws – a preliminary note.' Report of the Third International Conference 1906 on Genetics : hybridisation (the cross-breeding of genera or species), the cross-breeding of varieties, and general plant-breeding. London: Royal Horticultural Society (cited on page 26).

Yule, G. U. (1907). On the Theory of Inheritance of Quantitatively Compound Characters on the Basis of Mendel's Laws. *Biometrika* **5**: 481–482. DOI: 10.2307/2331701 (cited on page 19).

Zhang, Y., D. Li, and B. Sun (2015). Do Housekeeping Genes Exist? *PLOS ONE* **10**: e0123691. DOI: 10.1371/journal.pone.0123691 (cited on page 11).

Zheng, H. and W. Xie (2019). The role of 3D genome organization in development and cell differentiation. *Nature Reviews Molecular Cell Biology* **20**: 535–550. DOI: 10.1038/s41580-019-0132-4 (cited on page 5).

Zhu, J., F. He, S. Song, J. Wang, and J. Yu (2008). How many human genes can be defined as housekeeping with current expression data? *BMC Genomics* **9**: 172. DOI: 10.1186/1471-2164-9-172 (cited on page 11).