



# Possible mathematical definitions of the biological term “breed”

Gregor Langer

Gierkeweg 1, 4040 Linz, Austria

**Correspondence:** Gregor Langer (gregor.langer@gmx.net)

Received: 12 July 2017 – Revised: 23 May 2018 – Accepted: 30 May 2018 – Published: 14 June 2018

**Abstract.** For scientific discussions it is necessary to clearly define the terms used. Otherwise, scientific statements are open to interpretation and remain unclear which hampers scientific progress. A clear specification of scientific terms can be reached using tools such as mathematical definitions. So far, the biological term “breed” lacks a mathematical definition. Based on common biological literature, four mathematical definitions of the term “breed” for gonochoric species are proposed in this paper. These definitions were tested on the dog breed “whippet”. It was discussed if bully whippets and their offspring can be considered to be part of the whippet dog breed. Although all mathematical definitions proposed are consistent with common word-based definitions, the results of the bully whippet test differ. According to two definitions bully whippets may be considered part of the whippet breed. However, according to the other definitions bully whippets are not part of the whippet breed. Under certain conditions the introduced mathematical structure may also be used to define other biological terms like “variety” or “species”.

## 1 Introduction

The concept of breed is widely applied to many species (e.g., FAO, 2007), such as horses, cattle, sheep, or dogs. Although the breeding of animals was popular during the 19th century, the term “breed” was still not clearly defined at the beginning of the 20th century (Lloyd-Jones, 1915) and currently still remains unclear (e.g., Baumung et al., 2006 and FAO, 2012). The Federation Cynologique Internationale (FCI) defines the term “breed” as “A population of dogs sharing definable and inheritable phenotypic characteristics, after breeding undertaken by humans over a given period of time, allowing this population to be distinguished from other defined populations of dogs/or defined breeds.” (FCI, 2015) A definition not restricted to dogs, which is accepted by the Food and Agriculture Organization of the United Nations (FAO), can be found in FAO (2000); Galal (2005): “A subspecific group of domestic livestock with definable and identifiable external characteristics that enable it to be separated by visual appraisal from other similarly defined groups within the same species, or a group for which geographical and/or cultural separation from phenotypically similar groups has led to acceptance of its separate identity”. In Hall and Bradley (1995)

the term “breed” is defined as “a group of animals selected by man to have a uniform appearance that distinguishes them from other members of the same species”. All these definitions have the fact in common that a breed is a subspecific group of domestic livestock that share definable phenotypes and/or characteristics. Referring to the definitions, the phenotypes only have to allow an investigator (e.g., a breeder) to recognize the specific breed. Therefore, the definable phenotypes and/or characteristics can be considered to be arbitrary. According to the definitions given by Hall and Bradley (1995) and Galal (2005) it is not explicitly required that the specific phenotypes are inheritable. However, it is somehow clear that the definable characteristics need to be potentially inheritable, because otherwise one could employ properties like hair-cut or underweight as breed criterion. Additionally, if the characteristics can not be inherited, the offspring is not likely to belong to the specific breed any more. In Table 1 the differences of the three abovementioned breed definitions plus the definition of the Cat Fanciers’ Association (CFA) are outlined. All definitions require that the individuals of a certain breed belong to the same species, i.e., can interbreed.

**Table 1.** This table outlines the differences of the four word-based breeding definitions. All definitions have the fact in common that all individuals of a breed must belong to the same species.

Reference	Definition	Characteristics	Recognition
FCI (2015)	A population of dogs sharing definable and inheritable phenotypic characteristics, after breeding undertaken by humans over a given period of time, allowing this population to be distinguished from other defined populations of dogs/or defined breeds.	Inheritable	
FAO (2000)	A subspecific group of domestic livestock with definable and identifiable external characteristics that enable it to be separated by visual appraisal from other similarly defined groups within the same species, or a group for which geographical and/or cultural separation from phenotypically similar groups has led to acceptance of its separate identity.	Not explicitly inheritable	Visual
CFA (2018)	A breed is a group of domestic cats (subspecies <i>Felis catus</i> ) that the governing body of CFA has agreed to recognize as such.	Not explicitly inheritable	
Hall and Bradley (1995)	A group of animals selected by man to have a uniform appearance that distinguishes them from other members of the same species.	Not explicitly inheritable	Visual

In a similar fashion to the term “breed” for domestic animals, within the literature the terms “cultivar”, “race”, “variety”, and “subspecies” define specific groups (subsets) of domestic plants, humans, and wild animals within the set of a certain species (Bailey, 1923; Montagu, 1962; Kapadia, 1963; Bhopal, 2004). In this context it is interesting to note that the CFA denotes cat breeds as subspecies of *Felis catus* (FAO, 2012; CFA, 2018). However, the terms “cultivar” (Bailey, 1923), “race” (Keita, 1993; Keita et al., 2004; Mayr, 2002; Patten, 2009), “variety” (Kapadia, 1963), and “subspecies” (Mayr, 1982; Patten, 2010, 2015) also do not have generally accepted definitions and are, therefore, still under debate. Mayr (2003) writes that even the term “species” is still under debate. This so-called “species problem” is reviewed in papers such as de Queiroz (2005).

Additionally, all definitions proposed for these terms have been given in words up until now. Words are open to interpretation (Cain, 1953), and their meanings change over time. During the past few years science has attempted to describe “breeds” through genetic variation (e.g., Wiener et al., 2004; Lacy, 1995) and “subspecies” with the help of the so-called “75 % rule” (Amadon, 1949; Baker et al., 2002; Patten and Unitt, 2002), which already resembles a mathematical definition of these terms. According to Amadon (1949) the “75 % rule” defines a new subspecies if 75 % of a population are separable from all overlapping populations in a single characteristic. However, these methods also remain controversial as long as the terms “breed” and “subspecies” themselves are not properly defined (Patten, 2015). The lack of a clear definition of the term “breed” is not only an academic problem; for example, this issue leads to the fact that different breeding organizations use different classification criteria and, therefore, recognize different numbers of dog breeds (Gonzalez et al., 2014).

In the present paper, a special mathematical formalism helps to describe sexual reproduction. Mathematical defini-

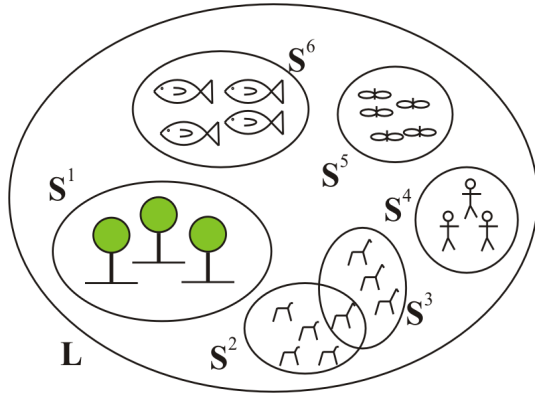
tions of the term “breed” are introduced with the help of this formalism. The mathematical definitions are based on the word-based definitions listed in Table 1. The definitions are meant to help organizations related to breeding (e.g., the FCI) to define or refine breeding standards and to judge whether existing or proposed breeds fulfill the required criteria. The proposed mathematical definitions were tested on the whippet dog breed (FCI standard 162, 2007). Whippets are racing dogs and within this breed an autosomal recessive genetic defect is very common (Mosher et al., 2007; Shadan, 2007). This genetic defect leads to massive muscle growth and the appearance of the dog then differs from the breeding ideal. Therefore, it is questionable if dogs with this appearance can be considered to be whippets. It is found that different mathematical definitions lead to different results. At the end of this paper the definitions proposed and their results are summarized and discussed. Furthermore, it is discussed if other biological terms, such as “variety”, “subspecies” or “species”, can be defined with the introduced mathematical formalism.

For readers who are not familiar with mathematical symbols, some of the mathematical symbols used in this paper are explained in the appendix.

## 2 Material and methods

### 2.1 Mathematical introduction

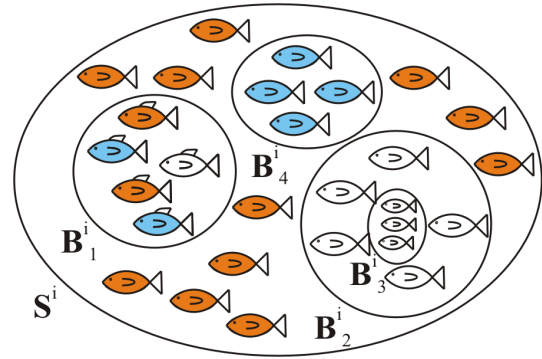
In the following, a mathematical structure is introduced that helps to describe the sexual reproduction between two individuals. For this purpose the set  $\mathbf{L}$  is introduced, which is the set of all possible creatures with sexual reproduction (see Fig. 1). “All possible” refers to existent, extinct, future, realized, and never realized but principally possible. The  $x$ th element of this set will be called  $l_x$ . The union of all species with sexual reproduction is a subset of  $\mathbf{L}$  (see Eq. 1). It is just a subset, because not all creatures need to belong to a



**Figure 1.**  $L$  is the set of all possible life-forms.  $S^i$  are the sets of species.

species. If for some reason it turns out that all creatures belong to a species the term “subset” includes equality anyhow. Species  $i$  will be denoted by  $S^i$ , e.g.,  $S^{\text{hazel}}$ ,  $S^{\text{dog}}$ ,  $S^{\text{lion}}$  and so on. In the following natural numbers are used, starting with 1, instead of words; hence,  $i \in \mathbb{N}$ . Here,  $S^1$  represents the set of all hazels,  $S^2$  that of all dogs, and so on.

Within this paper, only gonochoric reproduction with two sexes will be considered. In gonochoric reproduction, offspring is created via sex between a fertile male individual and a fertile female individual. Therefore, the subsets  $Lm$ ,  $Sm^i$  and  $Lf$ ,  $Sf^i$  are introduced, meaning fertile male and fertile female individuals (for multi-sexual reproduction, such a set would have to be introduced for every sex). The set  $Lm$  contains all possible fertile male creatures and  $Sm^i$  all possible fertile male individuals belonging to a certain species  $i$ . The same respective designation is used for the females. Additionally, infertile individuals are considered via  $Li$  and  $Si^i$ . It is important to note that only individuals that do not have the potential to become fertile (i.e., they are necessarily infertile during their whole life) belong to the sets  $Li$  or  $Si^i$  (e.g., individuals which lack gonads). For the sake of simplicity, it is additionally assumed that every individual has a single sex during its entire life (i.e., no switching between fertile male and fertile female or vice versa is possible). This means that a single individual can be an element of  $Sm^i$ ,  $Sf^i$ , or  $Si^i$ . The  $x$ th element of the corresponding sets will be called  $s_x^i$ ,  $sm_x^i$ ,  $sf_x^i$ , and  $si_x^i$ , for example. Therefore, the following relations hold:



**Figure 2.** A set of individuals belonging to species  $S^i$  can be divided into subsets  $B_j^i$ . All individuals belonging to a certain subset  $B_j^i$  belong to the same breed.

$$\begin{aligned}
 L &= Lm \cup Lf \cup Li \supseteq S^1 \cup S^2 \cup \dots \cup S^n \\
 Lm &\supseteq Sm^1 \cup Sm^2 \cup \dots \cup Sm^n \\
 Lf &\supseteq Sf^1 \cup Sf^2 \cup \dots \cup Sf^n \\
 Li &\supseteq Si^1 \cup Si^2 \cup \dots \cup Si^n \\
 S^i &= Sm^i \cup Sf^i \cup Si^i \\
 L &= \{l_1, l_2, l_3, \dots, l_h\} \\
 S^i &= \{s_1^i, s_2^i, s_3^i, \dots, s_{g_i}^i\} \\
 Sm^i &= \{sm_1^i, sm_2^i, sm_3^i, \dots, sm_{k_i}^i\} \\
 Sf^i &= \{sf_1^i, sf_2^i, sf_3^i, \dots, sf_{l_i}^i\} \\
 Si^i &= \{si_1^i, si_2^i, si_3^i, \dots, si_{m_i}^i\}
 \end{aligned} \tag{1}$$

In Eq. (1),  $n \in \mathbb{N}$  is the number of all possible species,  $h \in \mathbb{N}$  is the number of all possible life-forms,  $g_i \in \mathbb{N}$  is the number of all possible individuals belonging to species  $i$ ,  $k_i \in \mathbb{N}$  is the number of all possible fertile male individuals belonging to species  $i$ ,  $l_i \in \mathbb{N}$  is the number of all possible fertile female individuals belonging to species  $i$ , and  $m_i \in \mathbb{N}$  is the number of all possible sterile individuals belonging to species  $i$ . The number of elements of set  $S^i$  is equal to all elements that belong to the union of  $Sm^i$ ,  $Sf^i$ , and  $Si^i$ . Therefore,  $g_i = k_i + l_i + m_i$ .

From the introduction we know that a breed is a subset of a certain species. For the species  $S^i$ ,  $i \in \mathbb{N}$ , breeds are denoted by  $B_j^i$ ,  $j \in \mathbb{N}$  (see Fig. 2 and Eq. 2). The superscript  $i$  denotes the species (e.g., dog) and the subscript  $j$  denotes the specific breed (e.g., whippet). Here, a distinction is also made between fertile male, fertile female, and sterile individuals. Therefore, we can write

$$\begin{aligned}
 S^i &\supseteq B_1^i \cup B_2^i \cup \dots \cup B_{q_i}^i \\
 B_j^i &= Bm_j^i \cup Bf_j^i \cup Bsi_j^i
 \end{aligned} \tag{2}$$

In Eq. (2),  $q_i \in \mathbb{N}$  is the number of all breeds belonging to species  $i$ . Here, it is clear that the union of all breed sets is just a subset of the corresponding species set, because it is evident that e.g., not all dogs belong to a breed. The breed set  $\mathbf{B}_j^i$  consists of the union of all fertile males, fertile females and infertile individuals that fulfill the corresponding conditions. In principle, it is possible that a certain individual from species  $i$  is an element of several breed sets (compare with Fig. 2).

In order to consider the process of gonochoric reproduction and to formulate the term “breed” via a mathematical equation, two operators, i.e., “ $\circ$ ” and “ $\bullet$ ”, are introduced. Commonly, in mathematics an operator, e.g., “+”, combines two elements and gives a third element, e.g.,  $a + b = c$ . To describe the process of gonochoric reproduction of two individuals such a structure is insufficient, as gonochoric parents can potentially have a large number of offspring; furthermore each offspring may be different. Therefore, an operator is needed that combines two elements of the set  $\mathbf{L}$ , e.g.,  $l_x \circ l_y$ , and the result is the set of all possible offspring between these two elements, such as  $\{l_{x,y,1}, l_{x,y,2}, l_{x,y,3}\}$ . Additionally, we define an operator which also gives the corresponding set of probabilities for each offspring to be realized:

$$\begin{aligned} l_x \circ l_y &= \mathbf{L}_{x,y} \\ l_x \bullet l_y &= \mathbf{P}_{x,y} \end{aligned} \quad (3)$$

The set  $\mathbf{L}_{x,y}$  consists of all possible offspring between an individual  $l_x$  and an individual  $l_y$ :  $\mathbf{L}_{x,y} = \{l_{x,y,1}, l_{x,y,2}, \dots\}$ . The set  $\mathbf{P}_{x,y}$  consists of pairs  $(l_{x,y,i}, p_{x,y,i})$ , where  $p_{x,y,i}$  is the corresponding probability value, i.e., how probable it is that offspring  $l_{x,y,i}$  will be realized. The probability is element of the rational numbers, i.e.,  $p_{x,y,i} \in \mathbb{Q}$ . This means that offspring  $l_{x,y,1}$  is realized with a probability of  $p_{x,y,1}$  and so on. Therefore, the set  $\mathbf{P}_{x,y}$  takes the following form:  $\mathbf{P}_{x,y} = \{(l_{x,y,1}, p_{x,y,1}), (l_{x,y,2}, p_{x,y,2}), \dots\}$ . The elements as well as the corresponding probability values have to be determined according to laws of heredity. How to determine the elements of  $\mathbf{L}_{x,y}$  and  $\mathbf{P}_{x,y}$  in practice is shown by a simple example in Sect. 4.

$(\mathbf{L}, \circ, \bullet)$  is a mathematical structure. The operator  $\circ$  maps two elements of set  $\mathbf{L}$  onto the set of possible offspring (see Eq. 4). Let us name this set  $\mathbf{O}$ , for offspring. It contains the sets for all possible offspring combinations of the elements of  $\mathbf{L}$ :  $\mathbf{O} = \{\dots, \mathbf{L}_{x,y}, \mathbf{L}_{x,y+1}, \dots\} = \{\{l_1\}, \dots, \{l_1, l_2\}, \dots, \{l_1, l_2, l_3\}, \dots\}$ . Hence, it contains sets with single elements, with two elements, with three elements, and so on in all possible combinations. The operator  $\bullet$  maps two elements of set  $\mathbf{L}$  onto the set  $\mathbf{P} = \{\dots, \mathbf{P}_{x,y}, \mathbf{P}_{x,y+1}, \dots\} = \{\{(l_1, p_1)\}, \dots, \{(l_1, p_1), (l_2, p_2)\}, \dots\}$ , which contains doublets of the form  $\mathbf{L} \times \mathbb{Q}$  (see Eq. 4).

$$\begin{aligned} \circ : \mathbf{L} \times \mathbf{L} &\rightarrow \mathbf{O} \\ \bullet : \mathbf{L} \times \mathbf{L} &\rightarrow \mathbf{P} \end{aligned} \quad (4)$$

Both operators are commutative. This means that  $l_x \circ l_y = l_y \circ l_x$  and  $l_x \bullet l_y = l_y \bullet l_x$ . A notation like  $l_x \circ (l_y \circ l_z)$  makes no sense as the term in brackets is a set and both operators can only be applied to single elements of  $\mathbf{L}$ . Therefore, the introduced structure is not associative, which is typical for structures that describe inheritance, e.g., genetic algebras (e.g., Bernstein, 1942; Etherington, 1939; Reed, 1997; Ladra and Rozikov, 2013). The introduced structure only allows for the combination of two elements of the set  $\mathbf{L}$ . There is no neutral or inverse element. A neutral element fulfills the equation  $n \circ l = l$ . However, even for asexual cloning the equation looks like  $l \circ l = \{l\}$ . Therefore, from a mathematical point of view the introduced structure is boring. However, it may be used to give a set-theoretic definition of the term “breed”.

For most combinations of  $l_x$  and  $l_y$  the resulting sets,  $\mathbf{L}_{x,y}$  and  $\mathbf{P}_{x,y}$ , will be empty sets. For example, in the case where  $l_x$  is an element of  $\mathbf{Lm}$  and  $l_y$  is an element of  $\mathbf{Lf}$ , but  $l_x$  and  $l_y$  do not belong to the same species. (In common definitions of the term “species”, such as Mayr, 2003, individuals of different species may produce (infertile) offspring. For example horses and donkeys, or lions and tigers. For simplicity, in this paper the term “species” is used in such a way that it obeys the first row of Eq. (5). This stands in contradiction to common definitions based on words. However, this will not have any influence on the mathematical definition of the term “breed”. In principle one could name the sets that obey Eq. (5) with a new term, e.g., “hyperspecies”. However, as long as the terms “species” and “subspecies” lack mathematical definitions it probably makes no sense to introduce new terms.) In the following, we consider elements  $l_x$  that are elements of some set  $\mathbf{S}^i$  and will write  $\mathbf{sm}_x^i$ ,  $\mathbf{sf}_x^i$ , or  $\mathbf{si}_x^i$  instead. In Eq. (5) all combinations that result in the empty set are listed:

$$\begin{aligned} \mathbf{sm}_x^i \circ \mathbf{sf}_y^j &= \{\}, \quad \text{if } i \neq j \\ \mathbf{sm}_x^i \circ \mathbf{sm}_y^j &= \{\}, \quad \forall i, j \\ \mathbf{sf}_x^i \circ \mathbf{sf}_y^j &= \{\}, \quad \forall i, j \\ \mathbf{si}_x^i \circ \mathbf{si}_y^j &= \{\}, \quad \forall i, j \\ \mathbf{sm}_x^i \circ \mathbf{si}_y^j &= \{\}, \quad \forall i, j \\ \mathbf{sf}_x^i \circ \mathbf{si}_y^j &= \{\}, \quad \forall i, j \end{aligned} \quad (5)$$

Equation (5) expresses the following: the first row states that fertile males and fertile females belonging to different species can not produce any offspring. The following rows state that two fertile male individuals, two fertile female individuals, two infertile individuals, etc. can not have any offspring, even if they are from the same species.

Only in cases where the operator “ $\circ$ ” is applied to individuals between  $\mathbf{Sm}^i$  and  $\mathbf{Sf}^i$ , does a set unequal to the empty set result:

$$\begin{aligned}
 sm_x^i \circ sf_y^i &= \mathbf{Sm}_{x,y}^i \cup \mathbf{Sf}_{x,y}^i \cup \mathbf{Si}_{x,y}^i = \mathbf{S}_{x,y}^i \\
 \mathbf{S}_{x,y}^i &= \left\{ s_{x,y,1}^i, s_{x,y,2}^i, s_{x,y,3}^i, \dots, s_{x,y,a}^i, \dots, s_{x,y,w_{x,y,i}}^i \right\} \quad (6) \\
 \mathbf{Sm}_{x,y}^i &= \left\{ sm_{x,y,1}^i, sm_{x,y,2}^i, \dots, sm_{x,y,a}^i, \dots, sm_{x,y,r_{x,y,i}}^i \right\} \\
 \mathbf{Sf}_{x,y}^i &= \left\{ sf_{x,y,1}^i, sf_{x,y,2}^i, \dots, sf_{x,y,a}^i, \dots, sf_{x,y,t_{x,y,i}}^i \right\} \\
 \mathbf{Si}_{x,y}^i &= \left\{ si_{x,y,1}^i, si_{x,y,2}^i, \dots, si_{x,y,a}^i, \dots, si_{x,y,o_{x,y,i}}^i \right\}
 \end{aligned}$$

In Eq. (6) the genotype of the fertile male individual  $sm_x^i$  and the fertile female individual  $sf_y^i$  determine all possible genotypes of the common offspring, e.g.,  $s_{x,y,1}^i$ . The number of all possible offspring between the two individuals  $sm_x^i$  and  $sm_y^i$  is  $w_{x,y,i} \in \mathbb{N}$ . The sum of  $|\mathbf{Sm}_{x,y}^i|$  plus  $|\mathbf{Sf}_{x,y}^i|$  plus  $|\mathbf{Si}_{x,y}^i|$  is equal to  $|\mathbf{S}_{x,y}^i|$ . Therefore:  $r_{x,y,i} + t_{x,y,i} + o_{x,y,i} = w_{x,y,i}$ . For completeness, it is noted that not all elements of  $\mathbf{S}_{x,y}^i$  are necessarily elements of  $\mathbf{S}^i$ . Otherwise, it would not be possible for new species to arise over time. All elements of  $\mathbf{S}_{x,y}^i$  will be realized e.g., if two individuals have a high number of offspring. In reality often only a few elements of the set  $\mathbf{S}_{x,y}^i$  will be realized. In Eq. (6), the set  $\mathbf{S}_{x,y}^i$  can only be determined with knowledge of the genotypes of  $sm_x^i$  and  $sf_y^i$ . Nevertheless, in the following the single elements of  $\mathbf{S}^i$ ,  $s_x^i$  and  $s_y^i$ , will be characterized via a vector of phenotypes and characteristics, e.g., height, color of the nose, length of fur, and so on (Eq. 7). The reason for this is that we want to give a definition of the term “breed”. From the introduction it is known that individuals belonging to a certain breed need to be recognizable by their phenotypes, e.g., appearance. If the genotype of the male,  $sm_x^i$ , and female,  $sf_y^i$ , individuals are known (e.g., genotype for height, color of fur, length of fur, and so on) then the offspring’s genotype can in principle be determined. The offspring are the elements of the set  $\mathbf{S}_{x,y}^i$ . Armed with the information regarding whether certain genes are dominant or recessive one can determine the corresponding phenotypes. However, the environment may significantly influence the phenotypic expression. The complexity of the relationship between genotype and phenotype considering environmental influences was reviewed by studies such as Rutherford (2000). To find the set of offspring  $\mathbf{S}_{x,y}^i$  for two individuals  $sm_x^i$  and  $sf_y^i$ , i.e., to solve Eq. (6), may be very complicated. This depends on how accurately the phenotypes are specified.

The phenotype “color” can be given by an RGB value, “length” and “height” in centimeters, “weight” in kilograms, and so on. An element of the species  $\mathbf{S}^i$  is then characterized by

$$\begin{aligned}
 s_x^i &= (\text{height/cm} = 35, \text{color of nose/[RGB]} \\
 &= (0.3, 0.3, 0.8), \text{length of fur/cm} = 3, \dots, \text{sex} = \text{male}) \quad (7)
 \end{aligned}$$

If one defines that height is always given in centimeters and is always the first vector element, one can write

$$\begin{aligned}
 s_x^i &= (35, (0.3, 0.3, 0.8), 3, \dots, \text{male}) \\
 &= (s_x^i[1], s_x^i[2], s_x^i[3], \dots) \quad (8)
 \end{aligned}$$

In Eqs. (7) and (8) the last vector element specifies the sex. In cases where the individual  $s_x^i$  is a fertile male or fertile female this entry is “male” or “female”, respectively. Otherwise, the entry is “infertile”. For the sake of simplicity, in this paper the notations  $sm_x^i$ ,  $sf_y^i$ , and  $si_z^i$  are used.  $sm_x^i$  can be considered to be  $s_x^i$ , with “male” being the last vector entry.

The application of the operator “•” on the fertile male element  $sm_x^i$  and fertile female element  $sf_y^i$  results in a set which contains doublets of the offspring, i.e., the elements of  $\mathbf{S}_{x,y}^i$ , and the realization probability:

$$\begin{aligned}
 sm_x^i \bullet sf_y^i &= \mathbf{Pm}_{x,y}^i \cup \mathbf{Pf}_{x,y}^i \cup \mathbf{Pi}_{x,y}^i = \mathbf{P}_{x,y}^i \quad (9) \\
 \mathbf{P}_{x,y}^i &= \left\{ (s_{x,y,1}^i, p_{x,y,1}^i), (s_{x,y,2}^i, p_{x,y,2}^i), (s_{x,y,3}^i, p_{x,y,3}^i) \right. \\
 &\quad \left. \dots, (s_{x,y,a}^i, p_{x,y,a}^i), \dots, (s_{x,y,w_i}^i, p_{x,y,w_i}^i) \right\}
 \end{aligned}$$

The element  $s_{x,y,1}^i$ , is realized with the probability of  $p_{x,y,1}^i$  and so on. The corresponding probability values can be determined by means of (epi)genetics plus the consideration of the environmental influence, if the environment has an influence on the genetic expression. In Sect. 4, how to solve Eqs. (6) and (9) is demonstrated by way of a simple example.

## 2.2 Breeding standard

Let us recapitulate: in the previous section, we introduced a set of possible life-forms with sexual reproduction  $\mathbf{L}$ , sets of species  $\mathbf{S}^i$ , and sets of breeds  $\mathbf{B}_j^i$ . Additionally, the operator “o” which symbolizes an act of reproduction between two individuals and gives the set of all possible offspring and the operator “•” that additionally gives the corresponding probability values were introduced. Let us now focus on the sets  $\mathbf{B}_j^i$ .

From the introduction it is known that if a certain individual belongs to a certain set  $\mathbf{B}_j^i$  it needs to possess certain phenotypes or characteristics. These phenotypes (e.g., height, length of fur) and characteristics (e.g., aggressiveness) can be arbitrarily chosen by the investigator (e.g., the breeder). We now introduce the breeding standard set  $\mathbf{B}_j^i$  as the set of individuals that fulfill certain requirements:

$$\mathbf{B}_j^i = \left\{ s_x^i \in \mathbf{S}^i \mid a_1 \leq s_x^i[1] \leq a_2, b_1 \leq s_x^i[2] \leq b_2, \dots \right\} \quad (10)$$

Equation (10) means that all individuals  $s_x^i$  that belong to a certain species  $i$  are members of  $\mathbf{B}_j^i$  if the vector elements lie within a defined interval. The first vector element  $s_x^i[1]$ , corresponding to height, has to lie in the interval  $[a_1, a_2]$ , and

so on. The vector elements correspond to phenotypic quantities (compare with Eqs. 7 and 8). Equation (10) considers neither the inheritance of the phenotype nor the uniform appearance of a group of individuals (compare with Table 1). Therefore, Eq. (10) is not a sufficient definition of the term “breed”. However, it can be seen later in this paper that the  $\mathbf{B}_j^i$  set can be used as a starting point for the definitions.

In order to be able to formulate the following definitions in a mathematical way, the set  $\mathbf{A}_{j,x,y}^i$  is introduced:

$$\mathbf{A}_{j,x,y}^i = \left\{ a \mid s_{x,y,a}^i \in \mathbf{B}_j^i \right\} \quad (11)$$

Equation (11) defines the set  $\mathbf{A}_{j,x,y}^i$ , which contains all indices  $a$  belonging to the elements of  $\mathbf{S}_{x,y}^i$  that fulfill the breeding standard, i.e., all common offspring of individual  $sm_x^i$  and individual  $sf_y^i$  which are part of set  $\mathbf{B}_j^i$ . The set  $\mathbf{A}_{j,x,y}^i$  is necessary in order to calculate the probability that offspring of two individuals  $sm_x^i$  and  $sf_y^i$  fulfill the breeding standard of breed  $j$ :

$$P_{j,x,y}^i = \sum_{a \in \mathbf{A}_{j,x,y}^i} P_{x,y,a}^i \quad (12)$$

In Eq. (12) we sum up all the probability values  $P_{x,y,a}^i$  of  $\mathbf{P}_{x,y}^i$  given by the set  $\mathbf{A}_{j,x,y}^i$ . Therefore,  $P_{j,x,y}^i$  gives the probability that two parents  $sm_x^i$  and  $sf_y^i$  produce an offspring that is part of the breeding standard set  $\mathbf{B}_j^i$ . The probability given in Eq. (12) can be observed by the breeder, e.g., if three out of four puppies from parents  $x$  and  $y$  fulfill the breeding standard set of breed  $j$ , then  $P_{j,x,y}^i$  is 0.75. In reality, a large sample number, i.e., many offspring, is necessary so that the observed probability approaches the theoretical value.

### 2.3 Breeding standard for whippets

In FCI (2007) the breeding standards for whippets (No. 162) are listed. The document gives a sketch of the dog’s appearance, and on the following pages the form of the head, the color of the fur, the height, and so on are described mainly by text. The only property that is defined by a number is the height at the withers. In the following, this will be given in centimeters. All other breeding standards are defined in words. For the mathematical definition of the term “breed” itself this is no problem as we will see later. However, the aim of this paper is to get rid of a subjective interpretation of the term “breed”. Therefore, from this point of view the breeding standards should also be given in a mathematically unique way. The color of the nose, coat, or ears, may be given via an RGB vector  $\mathbf{v} = (R, B, G)$ . The vector  $(0, 0, 0)$  corresponds to perfect black. Such perfect black does not occur in nature. Consequently, it might then be sufficient to define the word ‘black’ using an RGB vector which is smaller than 0.1 component wise, i.e.,  $(R < 0.1, G < 0.1, B < 0.1)$ . In the following, we will write this as  $\mathbf{v} < 0.1$ . The length of the

fur could be given in centimeters instead of using the word “short”. And instead of the description “fine hair” the hair thickness could be specified in  $\mu\text{m}$ . In the following, three conditions that can be addressed by numbers are chosen. All additional conditions are symbolized by “...”.

The breeding standard (FCI standard 162, 2007) gives us the breeding standard sets  $\mathbf{Bm}_{162}^2$  and  $\mathbf{Bf}_{162}^2$ . The subscript 162 denotes the breed whippet:

$$\mathbf{Bm}_{162}^2 := \left\{ sm^2 \in \mathbf{Sm}^2 \mid \text{height} \in [47, 51], \text{ color of nose} = \mathbf{v} < 0.1, \text{ length of fur} \in [l_{\min}, l_{\max}], \dots \right\} \quad (13)$$

$$\mathbf{Bf}_{162}^2 := \left\{ sf^2 \in \mathbf{Sf}^2 \mid \text{height} \in [44, 47], \text{ color of nose} = \mathbf{v} < 0.1, \text{ length of fur} \in [l_{\min}, l_{\max}], \dots \right\}$$

$$\mathbf{B}_{162}^2 = \mathbf{Bm}^2 \cup \mathbf{Bf}^2$$

Equation (13) states that an individual that is an element of a certain species – in this case a dog – belongs to the breeding standard set  $\mathbf{B}_{162}^2$  of the whippet dog breed, if it fulfills the criteria on the right side. The criteria are slightly different depending on the sex. Therefore, two different breeding standard sets, i.e.,  $\mathbf{Bm}_{162}^2$  and  $\mathbf{Bf}_{162}^2$ , are necessary. The criteria are man-made and arbitrary. Note that black is only the preferred nose color. For different colored coats different nose colors may be allowed. However, as a translation of all these requirements into a mathematical formulation would be rather lengthy only the preferred color is given in Eq. (13).

According to the breeding standard it is not clear whether infertile individuals are part of the whippet breed or not. On the last page of the breeding standard FCI (2007) it says “Only functionally and clinically healthy dogs, with breed typical conformation should be used for breeding”. On one hand, one can interpret this as general advice to the breeder. On the other hand, this sentence is part of the official breeding standard. Equation (13) is meant in such a way that all dogs that belong to the set  $\mathbf{B}_{162}^2$  can be used for breeding. If dogs are disqualified for breeding then they are not part of the breeding standard sets given in Eq. (13). Infertile individuals can definitely not be used for breeding. Therefore, in the following it is considered that infertile dogs are not part of the breeding standard set.

The requirements for the breeding standard set  $\mathbf{B}_j^i$  are not necessarily limited to physical characteristics and could also include characteristics like aggressiveness. The only requirement for the characteristics is that they can be described by a mathematical formalism. Mathematical formalisms are able to solve surprisingly complex problems. For example, it was shown by Liu et al. (2012) that it is possible to classify the breed of a dog by a photo of its face with the help of a mathematical algorithm implemented on a computer.

### 3 Results

In Eqs. (10) and (13) the breeding standard set  $\mathbf{B}'$  consists of all individuals with certain phenotypes. This is already close to the word-based definition of the term “breed” but does not consider that the phenotypes have to be inheritable or that the subspecific group has a uniform appearance (compare with Table 1). Therefore, the breeding standard alone, is considered to be insufficient as a definition of the term “breed”. In the following, inheritance is included and four possible definitions of the breed set  $\mathbf{B}'_j$  are given. In principle, an infinite number of possibilities exist regarding how to include the inheritance; therefore, the four examples are only meant to demonstrate some mathematical elements that may be part of a final definition.

#### 3.1 Definition 1

Naively, one would assume that offspring of two individuals that are recognized as belonging to a certain breed  $j$  (i.e., fulfill the breeding standard  $\mathbf{B}'^i_j$ ) belong to the same breed as their parents. Mathematically, this can be formulated as follows:

$$\mathbf{B}'_j := \left\{ s_{x,y,a} \in \mathbf{S}^i \mid s_{x,y,a} \in \left( \text{sm}_x^i \circ \text{sf}_y^i \right) \wedge \left( \text{sf}_y^i \in \mathbf{Bf}^i_j \right) \right. \\ \left. \wedge \left( \text{sm}_x^i \in \mathbf{Bm}^i_j \right) \wedge \left( \text{sm}_x^i \circ \text{sf}_y^i \right) \cap \mathbf{B}'^i_j \neq \{ \} \right\} \quad (14)$$

Equation (14) states that the breed consists of individuals  $s_{x,y,a}$  that belong to a certain species  $i$ , for which the requirements on the right side are fulfilled. Both, father ( $\text{sm}_x^i$ ) and mother ( $\text{sf}_y^i$ ), are elements of the breeding standard sets  $\mathbf{Bm}^i_j$  and  $\mathbf{Bf}^i_j$  respectively (compare with Eqs. 10 and 13). A possible offspring  $s_{x,y,a}$  is automatically an element of  $\mathbf{B}'_j$  and does not necessarily have to fulfill the breeding standard set  $\mathbf{B}'^i_j$ . In order to consider the “definable and inheritable phenotype” from the word-based definitions (FCI, 2015), the mathematical definition requires that at least some of the possible fertile offspring fulfill the criteria for the breeding standard, i.e.,  $(\text{sm}_x^i \circ \text{sf}_y^i) \cap \mathbf{B}'^i_j \neq \{ \}$ .

Let us apply Eq. (14) to the whippet dog breed. Dogs that fulfill the breeding standard for whippets may be part of the breed if their parents fulfill the breeding standard. However, according to Eq. (14) it does not matter what the offspring look like. Although a bully whippet does not obey the breeding standard – i.e., the conditions for  $\mathbf{B}'^i_{162}$  (Eq. 13) – it would nevertheless be a whippet if its parents obey the breeding standard. This however stands in contradiction to the word-based definition (Sect. 1), namely that a breed shares definable and inheritable phenotypic characteristics. However, what does “definable” and “inheritable” mean exactly? Is it sufficient that 90 % of the offspring obey the criteria for  $\mathbf{B}'^i_j$ ? (compare, for example, with the “75 % rule” for subspecies.) Due to genetic disorders we can exclude that 100 % of all offspring will fulfill the desired criteria for the breeding standard  $\mathbf{B}'^i_j$ . In principle, one may require that at least 90 % of

the offspring between  $\text{sm}_x^i$  and  $\text{sf}_y^i$  fulfill the breeding standard set  $\mathbf{B}'^i_j$ . Under this condition, Eq. (14) would then be expressed as follows:

$$\mathbf{B}'_j := \left\{ s_{x,y,a} \in \mathbf{S}^i \mid s_{x,y,a} \in \left( \text{sm}_x^i \circ \text{sf}_y^i \right) \wedge \left( \text{sf}_y^i \in \mathbf{Bf}^i_j \right) \right. \\ \left. \wedge \left( \text{sm}_x^i \in \mathbf{Bm}^i_j \right) \wedge P^i_{j,x,y} \geq 0.9 \right\} \quad (15)$$

But as 90 % is an arbitrary value, Eq. (14) only requires that the intersection of  $\mathbf{B}'$  with  $(\text{sm}_x^i \circ \text{sf}_y^i)$  must not be the empty set. Let us return to the whippets. From FCI standard 162 (2007) it can be concluded that bully whippets do not fulfill the breed criteria given in Eq. (13). Therefore, offspring of a bully whippet interbred with a normal whippet would not belong to the breed whippet according to Eq. (14). However, its offspring may lie within  $\mathbf{B}'^2_{162}$ . Thus, bully whippet’s offspring in the second generation may be whippets again, if both parents obey the breeding standard. In Eq. (14) even infertile individuals may be part of the breed if their parents fulfill the breeding standard.

#### 3.2 Definition 2

A breeder might desire a restrictive definition in which only an offspring that fulfills the breed criteria is part of the breed  $\mathbf{B}'_j$ :

$$\mathbf{B}'_j := \left\{ s_{x,y,a} \in \mathbf{S}^i \mid s_{x,y,a} \in \mathbf{B}'^i_j \cap \left( \text{sm}_x^i \circ \text{sf}_y^i \right) \right. \\ \left. \wedge \left( \text{sf}_y^i \in \mathbf{Bf}^i_j \right) \wedge \left( \text{sm}_x^i \in \mathbf{Bm}^i_j \right) \right\} \quad (16)$$

Firstly, a member of the breed  $s_{x,y,a}$  has to be part of the correct species  $\mathbf{S}^i$ . Furthermore, Eq. (16) states that common offspring of two elements of the breeding standard sets, i.e.,  $\text{sm}_x^i \in \mathbf{Bm}^i_j$  and  $\text{sf}_y^i \in \mathbf{Bf}^i_j$ , only belong to this breed if they fulfill the breeding standard criteria specified in  $\mathbf{B}'^i_j$ . This means that only fertile individuals belong to the breed, because infertile individuals are excluded from the breeding standard. It is also impossible to determine the breed of a dog by visual appraisal alone. In order to classify a dog it is necessary to have additional knowledge regarding the appearance of its parents. The indices  $x$  and  $y$  at  $s_{x,y,a}$  denote the ancestors. However, a situation may exist where  $s_{x,y,a} = s_{x+\alpha,y+\beta,a+\gamma}$ , with  $\alpha, \beta, \gamma \in \mathbb{N}_0$ , but only one of these elements is part of the breed set  $\mathbf{B}'^i_j$ . Hence, the fulfillment of the breeding standard  $\mathbf{B}'$  is not a sufficient criterium for being part of the breed set  $\mathbf{B}'^i_j$ ; for example, if one of the parents is not an element of the breeding standard set, e.g.,  $\text{sf}^i_{y+\beta} \notin \mathbf{Bf}^i_j$ , then the offspring  $s_{x+\alpha,y+\beta,a+\gamma}$  is not part of the breed, i.e.,  $s_{x+\alpha,y+\beta,a+\gamma} \notin \mathbf{B}'^i_j$ .

A bully whippet does not fulfill the breeding standard for whippets (FCI standard 162, 2007); therefore, a bully whippet can not be considered to be a whippet according to Eq. (16). The nomenclature bully whippet could already account for that fact. Offspring of a bully whippet is not a

whippet, because the bully whippet is not an element of the breeding standard set  $\mathbf{B}^i_j$ . Although, second generation bully whippet offspring may be whippets again in cases where both parents obey the breeding standard.

### 3.3 Definition 3

It is also possible to find a definition in which the parents' phenotype does not influence whether an individual belongs to a breed or not:

$$\mathbf{B}^i_j := \left\{ s^i_v \in \mathbf{S}^i \mid s^i_v \in \mathbf{B}^i_j \wedge \exists s^i_z \in \mathbf{B}^i_j : P^i_{j,v,z} \geq 0.9 \right\} \quad (17)$$

In Eq. (17), a member of the breed  $s^i_v$  logically has to belong to the corresponding species, i.e.,  $s^i_v \in \mathbf{S}^i$ . Further, members of a breed need to fulfill the breeding standard sets, i.e.,  $s^i_v \in \mathbf{B}^i_j$ . This excludes infertile individuals and bully whippets. In order to consider the “inheritable phenotype” from the word-based definitions, it is required that at least a single individual exists within the breeding standard set such that 90 % of the possible common offspring fulfill the breeding standard, i.e.,  $\exists s^i_z \in \mathbf{B}^i_j : P^i_{j,v,z} \geq 0.9$ .

Definition 3 is very similar to the pure breeding standard, with the exception that it includes requirements regarding inheritance. Therefore, according to Eq. (17), not all elements that fulfill the breeding standard are automatically elements of the breed set  $\mathbf{B}^i_j$ . An element that fulfills the breeding standard may only be part of the breed if an additional element  $s^i_z$  exists such that 90 % of their possible common offspring fulfill the breeding standard. In the following, breed definition 3 is applied on the bully whippet: A bully whippet does not fulfill the breeding standard for whippets. Therefore, a bully whippet is not part of the breed whippet. However, a bully whippet offspring may be part of the whippet breed if it only has a single allele for massive muscle growth. Bully whippets' offspring in second generation also might belong to the whippet breed.

### 3.4 Definition 4

A more complex definition may consist of several equations:

$$\begin{aligned} \mathbf{Bm}^i_j &:= \left\{ sm^i_v \mid sm^i_v \in \mathbf{Sm}^i \wedge \exists sf^i_z \in \mathbf{Bf}^i_j : P^i_{j,v,z} \geq 0.9 \right\} \\ \mathbf{Bf}^i_j &:= \left\{ sf^i_v \mid sf^i_v \in \mathbf{Sf}^i \wedge \exists sm^i_z \in \mathbf{Bm}^i_j : P^i_{j,v,z} \geq 0.9 \right\} \\ \mathbf{Bi}^i_j &:= \left\{ si_{x,y,a} \in \mathbf{Si}^i \mid si_{x,y,a} \in \left( sm^i_x \circ sf^i_y \right) \right. \\ &\quad \left. \wedge sm^i_x \in \mathbf{Bm}^i_j \wedge sf^i_y \in \mathbf{Bf}^i_j \right\} \\ \mathbf{B}^i_j &:= \mathbf{Bm}^i_j \cup \mathbf{Bf}^i_j \cup \mathbf{Bi}^i_j \end{aligned} \quad (18)$$

The last line of Eq. (18) states that the breed  $\mathbf{B}^i_j$  is the union of the breed sets  $\mathbf{Bm}^i_j$ ,  $\mathbf{Bf}^i_j$ , and  $\mathbf{Bi}^i_j$ . The set  $\mathbf{Bm}^i_j$  contains all male individuals of species  $\mathbf{S}^i$  for which a female  $sf^i_z$  exists that fulfills the breeding standard, meaning

that 90 % of the common offspring lie within the breeding standard set. The same can be said for the  $\mathbf{Bf}^i_j$  set, only that male has to be replaced by female and vice versa. The set  $\mathbf{Bi}^i_j$  includes the infertile individuals that are offspring of members of  $\mathbf{Bm}^i_j$  and  $\mathbf{Bf}^i_j$ . Therefore, infertile individuals that are offspring of two members of the breed belong to the breed too.

Let us apply definition 4. Common offspring of a dog that fulfills the breeding standard, i.e.,  $sm^i_v \in \mathbf{Bm}^i_j$ , and another dog that fulfills the breeding standard, will very likely also fulfill the breeding standard. Therefore, a dog which is an element of the breeding standard set, e.g., a dog that is recognized as a whippet, may easily fulfill the condition  $\exists sf^i_z \in \mathbf{Bf}^i_j : P^i_{j,v,z} \geq 0.9$ . This is due to the fact that the phenotype is strongly related to the genotype. Therefore, common offspring of a dog that does not fulfill the breeding standard, i.e.,  $sm^i_v \notin \mathbf{Bm}^i_j$ , and a dog that fulfills the breeding standard will have a low probability of fulfilling the breeding standard. Hence, the condition  $\exists sf^i_z \in \mathbf{Bf}^i_j : P^i_{j,v,z} \geq 0.9$  excludes most dogs that do not fulfill the breeding standard from the breed set.

A bully whippet has two alleles for massive muscle growth. However, common offspring of a bully whippet and a whippet can carry only a single allele for massive muscle growth. A dog such as this fulfills the breeding standard. Therefore, bully whippets may be part of the breed according to Eq. (18). As stated, offspring of bully whippets either have one or two alleles for massive muscle growth. Therefore, from these individuals, an individual may exist from the common offspring that is part of the breeding standard. This means a bully whippet's offspring may also belong to the breed.

Infertile offspring of two whippets belong to the breed according to Eq. (18). Here no further conditions regarding inheritance have to be fulfilled.

## 4 Discussion

In the introduction of this paper it was mentioned that the scientific community is aware of the inadequate definition for the term “breed”. Despite this awareness, no solutions to perfect the definition have been presented. In this paper, methods outlining how to achieve a unique definition were put forward. However, an infinite number of definitions is possible. In Eqs. (15) and (17) alone the number 0.9 could be replaced by any other number. Which number is adequate? Additionally, the final definition does not need to have the form of any of the four suggested definitions. Therefore the following questions need to be considered: which requirements have to be addressed? Should the phenotypes of the parents have an influence? Should infertile individuals be members of a breed? Which definition can be applied in daily life?

In principle, an exact definition does not need to be formulated in a mathematical expression. All definitions proposed



can be translated into words. This was carried out for all four definitions below the corresponding equation. However, the meaning of words might change over time. This fact becomes evident from simply reading an old book. Therefore, an exact definition described in words may change in meaning over time. A mathematical definition will remain unique.

The main advantage of a unique definition may lie in the field of zoological research. If scientists or breeding organizations want to classify certain breeds, it is an advantage to know exactly what one is talking about; a scientific discussion or decision is difficult if one scientist claims that a certain animal is part of a certain breed and another scientist is convinced of the opposite. The FCI and other breeding organizations constantly define new breeds. Within the FCI there is an ongoing discussion about the introduction of varieties (Denis, 2016) in order to reduce the introduction of new breeds. A mathematical definition of the term “breed” could help find a solution to this debate. Most importantly, it would also make it necessary to mathematically define the term “variety”. Then, a distinct difference between these two terms would exist, which would make it possible to decide whether a certain set of dogs should be classified as a “breed” or as a “variety”.

Note that the four proposed definitions were formulated for gonochoric species with two sexes, which reproduce in a similar fashion to mammals. In the present formulation, the definitions can not be applied to multisexual, parthenogenetic, and hermaphroditic species. In parthenogenetic species, for example, an offspring only possesses a single parent. In order to address this fact, the equations have to be reformulated correspondingly. However, the formalism in its present form is sufficient for most animal breeds, such as dogs, cats, cattle and birds.

In Table 2 the results of the four proposed mathematical definitions of the term “breed” applied on the dog breed whippet are summarized. All of these definitions can be considered to be consistent with common word-based definitions from the literature. Nevertheless, the result of the whippet test differs. In addition to the four proposed definitions the pure breeding standard from Eq. (13) is also listed for comparison. The pure breeding standard does not consider the inheritance of the phenotype, nor the uniform appearance of the subspecific group (compare with Table 1). One could argue that the phenotype is encoded in the genotype and is, therefore, inheritable. However, a single individual within a large group, which by chance fulfills the breeding standard, does not belong to the corresponding breed according to the word-based definitions as it lacks the uniform appearance (see Table 1), but it would belong according to the breeding standard. The “yes”, “maybe”, and “no” of Table 2 are meant in the following sense: does an arbitrary dog, e.g., a dog from the street, with a certain phenotype and characteristics belong to the breed according to the respective breed definition? Dogs that possess a phenotype and characteristics defined by the breeding standard are definitely part of the breeding stan-

dard set. Bully whippets, however, are not part of the breeding standard set. The reason for this is that they do not fulfill the required phenotypic characteristics. As the allele for bully whippets is inherited recessively, a bully whippet’s offspring may fulfill the requirements for the breeding standard set. Infertile individuals can not be used for breeding; therefore they are excluded from the breeding standard.

Many breed organizations require documented bloodlines in order to recognize a breed. Additionally, many people may consider infertile individuals as being part of a breed, if these individuals fulfill all other phenotypic criteria and characteristics. In definition 1 (see Eq. 14), both parents need to fulfill the breeding standard in order for the offspring to belong to the breed. Additionally, infertile individuals may belong to the breed, if both parents are elements of the breeding standard set. As knowledge about the ancestors, i.e., the parents, is necessary for the classification, a dog from the street can not be classified with certainty. From Table 2 it can be seen that definition 1 may also contain bully whippets if both parents are elements of the breeding standard set. But it does not contain bully whippets’ offspring, because bully whippets are not element of the breeding standard set. Hence, if one knows that a dog is the offspring of a bully whippet it is definitely not element of the whippet breed. Although an animal might fulfill the breeding standard, according to definition 1 it can only be classified with the knowledge about the ancestors.

A breeder might prefer a restrictive breed definition, such as definition 2 outlined in this paper. According to definition 2 (see Eq. 16), neither bully whippets nor their offspring nor infertile individuals are part of the breed. Only dogs that fulfill the breeding standard may be part of the breed. Also here, knowledge about a dog’s ancestry is necessary in order to judge whether a certain individual is part of the breed or not. Definition 2 is probably closest to how breeds are currently classified by breeding organizations, which require documented bloodlines; normally, not only a single generation, i.e., the parents, but several generations must be documented. In principle, this can be done in Eq. (16), by adding requirements for the parents, e.g.,  $\dots (sf_y^i \in \mathbf{Bf}_j^i \cap (sm_v^i \circ sf_w^i)) \wedge (sm_v^i \in \mathbf{Bm}_j^i) \wedge (sf_w^i \in \mathbf{Bf}_j^i) \dots$ . Here,  $sm_v^i$  and  $sf_w^i$  are grandparents of the individual under consideration. However, the drawback of such a definition is that it can only be applied if documented bloodlines exist. For certain breeds this might not be the case.

Definition 3 was introduced in order to show that the parents’ phenotype is not necessary for a breed definition. The breed set according to definition 3 (see Eq. 17) and the breeding standard set contain almost the same elements (see Table 2). The difference is that definition 3 contains requirements regarding inheritance, namely that at least 90 % of the possible offspring of two individuals from a breed have to fulfill the breeding standard. Therefore, an animal that is an element of the breeding standard set is not automatically part

of the breed. In definition 3 it is possible, in principle, to replace the  $\exists$  with  $\forall$ . Then, a common offspring of an animal  $s_v^i$  with any animal  $s_z^i \in \mathbf{B}_j^i$  has to fulfill  $P_{j,v,z}^i \geq 0.9$ , i.e., it has to be element of the breeding standard with a probability of at least 90 %. Bully whippet’s offspring could no longer fulfill this condition. In order to once again include bully whippet offspring, the limit, i.e., 0.9, would need to be decreased to at least 0.75.

Definition 4 is based on definition 3. The difference is that it consists of several equations. This case allows for the inclusion of the infertile individuals into the breed definition. According to definition 4, not only individuals that fulfill the breeding standard but also bully whippets and bully whippets’ offspring may be part of the breed. Additionally, infertile individuals may be part of the breed in case where both parents are members of the breed. Definition 4 is probably the closest to what a layman would intuitively call a “breed”. If one looks at two or three generations one can judge by visual appraisal if certain conditions regarding the inheritance are fulfilled. Two or three generations are commonly alive at a particular time. Therefore, a breeder can judge if a certain set of individuals belongs to the same breed by a single investigation. Additionally, one does not have to exclude infertile individuals or individuals with certain ancestors, e.g., bully whippets. According to definition 4, visual appraisal of individuals belonging to different generations can be enough to judge if a certain set of animals fulfills the conditions to belong to a breed. No genetic or fertility tests are necessary. For breeds without a documented bloodline, this kind of definition seems to be most accurate. Furthermore, in definition 4 it is possible, in principle, to exchange the  $\exists$  with a  $\forall$ , which makes the conditions more restrictive.

In the following, two simple examples demonstrate how to apply the mathematical operators introduced in this paper. Let us take a male and a female whippet, each with a single allele for massive muscle growth. For simplicity only this single gene will be considered. The male individual will be called  $sm_x^2$  and the female one  $sf_y^2$ . The set of possible offspring,

$$sm_x^2 \circ sf_y^2 = \mathbf{S}_{x,y}^2 = \{(\text{bully, male}), (\text{normal, male}), (\text{bully, female}), (\text{normal, female}), (\text{bully, infertile}), (\text{normal, infertile})\}, \quad (19)$$

contains six elements. In this case, the elements are vectors with two components only, i.e., a single property and the sex. The first element  $s_{x,y,1} = (\text{bully, male})$  and the second element  $s_{x,y,2} = (\text{normal, male})$ . As the respective gene is inherited autosomally the sets for fertile male, fertile female, and infertile individuals contain the same number of elements:  $\mathbf{Sm}_{x,y}^2 = \{(\text{bully, male}), (\text{normal, male})\}$ ,  $\mathbf{Sf}_{x,y}^2 = \{(\text{bully, female}), (\text{normal, female})\}$ , and  $\mathbf{Si}_{x,y}^2 = \{(\text{bully, infertile}), (\text{normal, infertile})\}$ . The set containing the probability val-

ues is

$$sm_x^2 \bullet sf_y^2 = \mathbf{P}_{x,y}^2 \approx \{((\text{bully, male}), 0.25/2), ((\text{normal, male}), 0.75/2), ((\text{bully, female}), 0.25/2), ((\text{normal, female}), 0.75/2), ((\text{bully, infertile}), 0), ((\text{normal, infertile}), 0)\}, \quad (20)$$

if one assumes that most offspring will become fertile. A bully whippet will be realized with a probability of 25 %, i.e., two times 0.25/2, and a normal whippet with a probability of 75 %.

The second example: if a bully whippet mates with a whippet without massive muscle growth allele, then the set of offspring,  $sm_x^2 \circ sf_y^2 = \mathbf{S}_{x,y}^2 = \{(\text{normal, male}), (\text{normal, female}), (\text{normal, infertile})\}$ , contains only three elements. The property “normal” will be realized with a probability of 100 %, if genetic mutations are neglected.

It may be surprising that in the above examples the set  $\mathbf{S}_{x,y}^2$  has so few elements, namely only six or three. The number of elements strongly depends on the number of considered phenotypes and how accurately the phenotypes are described. In the example above we considered just a single phenotype, namely the muscle growth. Regarding the accuracy we differentiated between “inside breeding standard”, i.e., “normal”, and “outside breeding standard”, i.e., “bully”. This could be carried out similarly with other quantities and characteristics such as height or color. The calculation is more difficult if the phenotype depends on more than a single gene, but for a breeder the phenotypic differentiation between “short”, i.e., outside the breeding standard, “normal”, i.e., inside the breeding standard, and “high”, i.e., again outside the breeding standard, or “dark”, “normal”, and “light” is sufficient. Hence, the number of elements in  $\mathbf{S}_{x,y}^2$  will still be large if several phenotypes are considered simultaneously. Although, it should be possible to calculate them and the corresponding probabilities using a computer.

In the present paper, the breeding standard set was defined by phenotypic characteristics. The reason for this is that this is how breeds are defined e.g., by the FCI. A definition given by genotypes would have been simpler, as then the probability values of the elements of  $\mathbf{S}_{x,y}^2$  (see Eq. 6) can be determined without considering environmental influences. Currently, in many cases, the environmental influence on the phenotype can only be roughly estimated. Nevertheless, genotype and phenotype are related to each other. Therefore, breeds can be reliably recognized not only by means of phenotypes (e.g., in Liu et al., 2012 or Bradshaw and Goodwin, 1998) but also by means of genetic cluster analysis (e.g., Rosenberg et al., 2001). This shows that a genetic definition is possible in principle. However, from a historic point of view, one has to keep in mind that the concept of breed is older than the concept of genes. Additionally, it would be problematic to define breeds through genetic standards. The breeders would then have to constantly perform genetic tests on their breed, which would raise costs. Additionally, breed-

**Table 2.** This table provides a summary of the elements that are part of the breed sets according to the different breed definitions, if they are applied to the dog breed “whippet”. A “No” within the table means that the object is not part of the corresponding set (e.g., a bully whippet is definitely not part of the whippet’s breeding standard). A “Yes” within the table means that the object is definitely part of the corresponding set (e.g., dogs that fulfill the breeding standard are part of the breeding standard set). A “Maybe” means that the object may be part of the corresponding set if certain conditions are fulfilled (e.g., in definition 3: dogs that fulfill the breeding standard are only an element of the breeding standard set if another dog exists that fulfills the breeding standard, and the probability for common offspring being an element of the breeding standard exceeds 90 %).

	Element of breeding standard (whippet)	Bully whippet	Bully whippet’s offspring	Infertile individuals
Breeding standard (Eq. 13)	Yes	No	Maybe	No
Definition 1 (Eq. 14)	Maybe	Maybe	No	Maybe
Definition 2 (Eq. 16)	Maybe	No	No	No
Definition 3 (Eq. 17)	Maybe	No	Maybe	No
Definition 4 (Eq. 18)	Maybe	Maybe	Maybe	Maybe

ers would have to become experts in molecular biology in order to apply and interpret the tests correctly. Nevertheless, in order to combat diseases canine clubs offer or even require genetic tests (Farrell et al., 2015), e.g., the FCI is now requiring a genetic study for assessing the degree of genetic originality of a candidate breed for international breed recognition (Leroy and Trainin, 2018). Despite this fact, it is uncertain if genetic tests should be made mandatory for breeders (Leroy and Trainin, 2018; Farrell et al., 2015). Finally, the breeder or a customer is primarily interested in the phenotypes such as milk quantity or coat color and not in the underlying genotypes. Therefore, for practical reasons it may be reasonable to stay with the present situation, i.e., that the breeding standard is given by phenotypes. However, the mathematical definitions proposed in this paper also work, if the term “breed” becomes defined via genotypes. Instead of certain phenotypes, an individual of a certain breed needs to possess certain genotypes. This can be addressed in the same way as shown in Eq. (13). Instead of the phenotypes in Eq. (13) the corresponding genes are listed. An eventual breed definition can remain unchanged, because the definition itself only refers to the breeding standard set  $B'$ . Therefore, it is possible to find a unique definition of the term breed, using both today’s breeding standards and a future adaption of the underlying breeding standard to modern methods (e.g., genetic analysis), which would not require a redefinition of the term breed.

Finally, the mathematical structure itself is not necessarily limited to the biological term “breed”. From a scientific point of view it would be of interest to mathematically define the terms “subspecies”, “variety”, and “species”. However, the word-based definitions of the term “subspecies” include geographic information (e.g., Patten, 2010; Remsen, 2010). In the case of breeds, a whippet does not become a new breed just because it is sold to a different continent or is fed with different food: a whippet stays a whippet, regardless of if it lives in America or Europe or if it is exclusively fed with pork or beef. For wild living animals, however, the situation is different. In many common definitions of the term “sub-

species” (see e.g., Patten, 2010; Remsen, 2010) geographic distinction is sufficient to cause two groups of animals of the same species belong to different subspecies. The mathematical formalism introduced in this paper does not consider any geographic information; therefore, it is insufficient for the term “subspecies”. However, it might be used for terms for which no geographic information is necessary, e.g., the term “species”.

## 5 Conclusions

Currently, a clear definition of the biological term “breed” does not exist. In this paper, a mathematical structure was introduced that allows for the description of sexual reproduction between animals. With the help of this mathematical structure four possible definitions of the term “breed” were proposed. All definitions were based on definitions taken from common literature. The mathematical definitions were tested on the whippet dog breed. Whippets are racing dogs, and in this breed a genetic disorder is common. This genetic disorder leads to massive muscle growth, and the appearance of the dog differs from the breeding ideal. Such dogs are called bully whippets. In the test, it was investigated if bully whippets belong to the whippet breed or not. In two definitions, bully whippets were part of the breed. In the other two definitions, bully whippets were not part of the whippet breed. Additionally, it was tested if bully whippets’ offspring or infertile dogs belong to the breed. In both cases two definitions included such individuals and the other two excluded them. Nevertheless, all these mathematical definitions can be considered to be consistent with common word-based definitions of the term “breed” found in the literature. This shows that current definitions of the term “breed” are not unique.

The mathematical structure introduced in this paper was used to formulate possible mathematical definitions of the term “breed” for gonochoric species. Adaptions to the presented equations are necessary, in order to obtain definitions for multisexual, parthenogenetic, or hermaphroditic species.

In the present paper, it was argued that an infinite number of possibilities for a unique definition of the term “breed” exist. A committee of breeders and biologists first need to decide if a unique definition of the term “breed” is desired or necessary and which criteria this definition should include.

This paper is meant to give the breeders and biologists new ideas regarding how to define the terms “breed”, “variety”, “subspecies”, or even “species”. Future work might put forward other mathematical relations to define these terms. And maybe, in a similar fashion to the definition of the astronomic term “planet”, one day there will be generally accepted definitions of the terms “breed”, “variety”, “subspecies”, and “species” in biology.

**Data availability.** No data sets were used in this article.

**Appendix A: mathematical symbols**

In the following, some mathematical symbols used in this paper are explained:

$\mathbb{N}$	Set of natural numbers. In this paper this is $\{1,2,3,\dots\}$ .
$\mathbb{N}_0$	Set of natural numbers including 0, i.e., $\{0,1,2,3,\dots\}$ .
$\mathbf{A} \cup \mathbf{B}$	Union of set <b>A</b> and <b>B</b> . It contains all elements of <b>A</b> and <b>B</b> .
$\mathbf{A} \cap \mathbf{B}$	Intersection of set <b>A</b> and <b>B</b> . It contains those elements that <b>A</b> and <b>B</b> have in common.
$\mathbf{A} \supseteq \mathbf{B}$	Set <b>A</b> is a superset of <b>B</b> . Every element of <b>B</b> is also an element of <b>A</b> .
$ \mathbf{L} $	Number of elements of set <b>L</b> . In Eq. (1) this was $h$ .
$\{\}$	Empty set. A set that contains no elements.
$b \in \mathbf{S}$	$b$ is an element of set <b>S</b> .
$b \notin \mathbf{S}$	$b$ is not an element of set <b>S</b> .
$:=$	Is defined as.
$\exists$	There exists.
$\forall i$	For all $i$ . The index $i$ can have any possible value.
$\wedge$	Logical “and”.
$\approx$	Approximately.

**Competing interests.** The author declares that he has no conflict of interest.

**Acknowledgements.** The author thanks Veronika Jungwirth for proofreading the manuscript, and Andreas Langer and Johannes Kofler for mathematical assistance and helpful discussions.

Edited by: Nina Melzer

Reviewed by: Saeid Ansari-Mahyari, Michael Patten, and two anonymous referees

## References

- Amadon, D.: The Seventy-Five Per Cent Rule for Subspecies, *The Condor*, 51, 250–258, 1949.
- Bailey, L. H.: *Gentes Herbarum*, vol. 1, Ithaca, NY, Bailey Hortorium, New York State College of Agriculture at Cornell University, 1923.
- Baker, A. N., Smith, A. N. H., and Pichler, F. B.: Geographical variation in Hector’s dolphin: Recognition of new subspecies of *Cephalorhynchus hectori*, *J. Roy. Soc. New Zeal.*, 32, 713–727, 2002.
- Baumung, R., Cubric-Curik, V., Schwend, K., Achmann, R., and Sölkner, J.: Genetic characterisation and breed assignment in Austrian sheep breeds using microsatellite marker information, *J. Anim. Breed. Genet.*, 123, 265–271, 2006.
- Bernstein, S.: Solution of a Mathematical Problem Connected with the Theory of Heredity, *Ann. Math. Stat.*, 13, 53–61, 1942.
- Bhopal, R.: Glossary of terms relating to ethnicity and race: for reflection and debate, *J. Epidemio Commun. Health*, 58, 441–445, 2004.
- Bradshaw, J. W. S. and Goodwin, D.: Determination of behavioural traits of pure-bred dogs using factor analysis and cluster analysis; a comparison of studies in the USA and UK, *Res. Vet. Sci.*, 66, 73–76, 1998.
- Cain, A. J.: Geography, Ecology and Coexistence in Relation to the Biological Definition of the Species, *Evolution*, 7, 76–83, 1953.
- CFA: Definition of a breed, The Cat Fancier’s Association, available at: <http://cfa.org/Breeds/BreedStandards/DefinitionofBreeds.aspx>, last access: 15 May 2018.
- de Queiroz, K.: Ernst Mayr and the modern concept of species, *P. Natl. Acad. Sci. USA*, 102, 6600–6607, 2005.
- Denis, B.: Reflections on the procedure for the recognition of new breeds by the FCI, *Federation Cynologique Internationale*, available at: <http://www.fci.be/en/FCI-Scientific-Commission-71.html>, last access: 12th November 2016.
- Etherington, I. M. H.: Genetic algebras, *Proc. Roy. Soc. Edinburgh*, 59, 242–258, 1939.
- FAO: World watch list for domestic animal diversity, edited by: Scherf, B. D., Rome, available at: [http://webpages.icav.up.pt/PTDC/CVT/099782/2008/ReferencesforExomicsproposal\[PTDC-CVT-099782-2008\]/PDFfilesreferences\\_max30\[Exomics\]/ScherfBD2000\\_WorldWatchListforDomesticAnimalDiveristy-3rdedition.pdf](http://webpages.icav.up.pt/PTDC/CVT/099782/2008/ReferencesforExomicsproposal[PTDC-CVT-099782-2008]/PDFfilesreferences_max30[Exomics]/ScherfBD2000_WorldWatchListforDomesticAnimalDiveristy-3rdedition.pdf) (last access: 15 May 2018), 2000.
- FAO: The State of the World’s Animal Genetic Resources for Food and Agriculture – in brief, edited by: Pilling, D. and Rischkowsky, B., Rome, available at: <http://www.fao.org/3/a-a1260e.pdf> (last access: 15 May 2018), 2007.
- FAO: Intergovernmental technical working group on animal genetic resources for food and agriculture, Rome, available at: <http://www.fao.org/docrep/meeting/026/me879e.pdf> (last access: 15 May 2018), 2012.
- Farrell, L. L., Schoenebeck, J. J., Wiener, P., Clements, D. N., and Summers, K. M.: The challenges of pedigree dog health: approaches to combating inherited disease, *Canine Genet. Epidemiol.*, 2, 1–14, 2015.
- FCI: FCI standard No. 162, available at: <http://www.fci.be/Nomenclature/Standards/162g10-en.pdf> (last access: 17 July 2016), 2007.
- FCI: FCI standing orders, available at: <http://www.fci.be/medias/FCI-REG-RGT-en-2009.pdf> (last access: 18 August 2016), 2015.
- Galal, S.: Biodiversity in goats, *Small Ruminant Res.*, 60, 75–81, 2005.
- González, A., Luque, M., Herrera, M., González, C., Angón, E., and Rodero, E.: Usefulness of discriminant analysis in the morphofunctional classification of Spanish dog breeds, *Arch. Anim. Breed.*, 57, 2, <https://doi.org/10.7482/0003-9438-57-002>, 2014.
- Hall, S. J. G. and Bradley, D. G.: Conserving livestock breed biodiversity, *TREE*, 10, 267–270, 1995.
- Kapadia, Z. J.: Varietas and Subspecies, a Suggestion Towards Greater Uniformity, *Taxon*, 12, 257–259, 1963.
- Keita, S. O. Y.: The subspecies concept in zoology and anthropology, *J. Black Studies*, 23, 416–445, 1993.
- Keita, S. O. Y., Kittles, R. A., Royal, C. D. M., Bonney, G. E., Furbert-Harris, P., Dunston, G. M., and Rotimi, C. N.: Conceptualizing human variation, *Nat. Genet.*, 36, 17–20, 2004.
- Lacy, R. C.: Clarification of Genetic Terms and Their Use in the Management of Captive Populations, *Zoo Biol.*, 14, 565–578, 1995.
- Ladra, M. and Rozikov, U. A.: Evolution algebra of a bisexual population, *J. Algebra*, 378, 153–172, 2013.
- Leroy, G. and Trainin, Z.: Dog breeding and molecular tools: uses and concerns, available at: <http://fci.be/medias/SCI-ART-SAN-OUT-MOL-LER-TRA-en-1707.pdf>, last access: 6 February 2018.
- Liu, J., Kanazawa, A., Jacobs, D., and Behumeur, P.: Dog Breed Classification Using Part Localization, in: *ECCV 2012, Part I, LNCS*, edited by: Fitzgibbon, A., Springer, Heidelberg, vol. 7572, 172–185, 2012.
- Lloyd-Jones, O.: What is a Breed?, *J. Hered.*, 6, 531–537, 1915.
- Mayr, E.: Of What Use Are Subspecies?, *The Auk*, 99, 593–595, 1982.
- Mayr, E.: *The Biology of Race and The Concept of Equality*, *Daedalus*, 131, 89–94, 2002.
- Mayr, E.: *Das ist Evolution*, C. Bertelsmann Verlag, München, 2003.
- Montagu, A.: The concept of race, *Am. Anthropol.*, 64, 919–928, 1962.
- Mosher, D. S., Quignon, P., Bustamante, C. D., Sutter, N. B., Mellers, C. S., Parker, H. G., and Ostrander, E. A.: A Mutation in the Myostatin Gene Increases Muscle Mass and Enhances Racing Performance in Heterozygote Dogs, *PLOS Genet.*, 3, 779–786, 2007.

- Patten, M. A.: “Subspecies” and “race” should not be used as synonyms, *Nature*, 457, p. 147, 2009.
- Patten, M. A.: Null Expectations in subspecies diagnosis, *Ornithol. Monogr.*, 67, 35–41, 2010.
- Patten, M. A.: Subspecies and the philosophy of science, *The Auk*, 132, 481–485, 2015.
- Patten, M. A. and Unitt, P.: Diagnosability versus mean differences of sage sparrow subspecies, *The Auk*, 119, 26–35, 2002.
- Reed, M. L.: Algebraic Structure of Genetic Inheritance, *B. Am. Math. Soc.*, 34, 107–130, 1997.
- Remsen, J. V.: Subspecies as a meaningful taxonomic rank in avian classification, *Ornithol. Monographs*, 67, 62–78, 2010.
- Rosenberg, N. A., Burke, T., Elo, K., Feldman, M. W., Freidlin, P. J., Groenen, M. A. M., Hillel, J., Mäki-Tanila, A., Tixier-Boichard, M., Vignal, A., Wimmers, K., and Weigend, S.: Empirical Evaluation of Genetic Clustering Methods Using Multilocus Genotypes From 20 Chicken Breeds, *Genetics*, 159, 699–713, 2001.
- Rutherford, S. L.: From genotype to phenotype: buffering mechanisms and the storage of genetic information, *BioEssays*, 22, 1095–1105, 2000.
- Shadan, S.: Run, whippet, run, *Nature*, 447, p. 275, 2007.
- Wiener, P., Burton, D., and Williams, J. L.: Breed relationships and definition in British cattle: a genetic analysis, *Heredity*, 93, 597–602, 2004.