E3S Web of Conferences **420**, 01001 (2023) https://doi.org/10.1051/e3sconf/202342001001
EBWFF 2023

Data distribution and asymmetry in leaf blade *Prunus padus* L.

Sergey Baranov[1,*], *Igor* Zykov[2], *Aleksandra* Prokopenko[2], *Lubov* Fedorova[3], and *Irina* Lipponen[4]

[1]Vladimir State University named after A.G. and N.G. Stoletov, Gor'kogo, 79, 600005, Russian Federation
[2]State Humanitarian Technological University, 22, Zelenaya Street, 142611 Orekhovo-Zuevo, Russian Federation
[3]I.M. Sechenov First Moscow State Medical University, 119991, 8-2 Trubetskaya str. Moscow, Russian Federation
[4] Polar-alpine botanical garden, Murmansk region, Kirovsk, 184209, Russian Federation

Abstract. To test the level of developmental stability of plants, an indicator of the asymmetry of their leaf blades is often used. The resulting systematic additive and multiplicative errors depend on the nature of the distribution sample data. On the example of bird cherry (*Prunus padus* L.) leaf blades, two types of distribution of sample data are demonstrated – normal and exponential. Statistical analyzes were performed in STATISTICA 10 (Stat.Ink). The exponential form of the distribution did not allow obtaining the mean value of the sample. The normalizing transformation of the sample contributed to obtaining the mean value and finding the relative level of the developmental stability of the population. In the Murmansk population (Apatity), a high variability of leaf blades was revealed, but, despite a significant asymmetry obtained in the normalizing difference formula, it did not have a statistically significant fluctuating asymmetry (FA). The sample from the Vladimir population (Vladimir) with the highest level of FA (0.018±0.004) was characterized by high values of kurtosis, skew and variance which was confirmed in a two-way analysis of variance (p <<0.05). In the Moscow population (Elektrostal) a high level of asymmetry was also obtained in the normalizing formula, but was not confirmed in a 2-way ANOVA. Thus, only the population from Vladimir showed a deviation in developmental stability in both methods. The authors recommend paying attention to the nature of the distribution of the difference between the left and right trait values, and using a two-way analysis of variance for the final determination of the FA value.

1 Introduction

Fluctuating asymmetry (FA) is a non-directional deviation from zero of the difference in the values of bilateral traits with its normal distribution. The fluctuation variability is common among symmetrical traits and can characterize the quantitative variability of

* Corresponding author: bar.serg58@gmail.com

biomass, color, and other parameters of the genotypic and phenotypic norm of the reaction. According to the value of the FA index, developmental stability judge as a deviation from the homeostasis of population development. The standard error of sample data is a common characteristic of the descriptive statistics of the FA index depending on its value and on the measurement error made by the researcher. A large error distorts the value with a decrease in the statistical significance of FA, which is often because of the presence of directional asymmetry (DA).

The selection of the most symmetrical leaves makes it possible to measure and compare such a trait as the width of their halves. Checking for directional asymmetry often shows its absence even with a sufficiently large sample size, for example, 50-100 leaves. High rates of kurtosis and skew show other types of asymmetry, including antisymmetry [1]. Numerically, the FA index is determined by the absolute value of the ratio of the difference in the values of traits to their sum.

Due to the large number of labels and random increase in the number of samples, the method of geometric morphometrics increases the number of degrees of freedom in the analysis of variance by several hundred times and normalizes the sample. As a result, the researcher gets a directional asymmetry, often exceeding the fluctuating one, for example, in terms of the mean square MS interaction sample × side [2]. Such a discouraging demonstration of DA makes it difficult to search for the level of developmental stability, confirming the multiplicity of manifestations of asymmetry in the sample.

The increasing number of measurements, photographs, and sample size reduces the measurement error. The advantage of Procrustes analysis with repeated measurements is undeniable. However, the question of how to eliminate directional asymmetry when comparing two or more samples remains unresolved.

The studies of representatives of the genus Padus have shown that plate's bilateral asymmetry is sensitive to automobile emissions [3], therefore, the search for additional indicative features [4] for a detailed analysis of the ecological status of the environment continues [5].

We have noted a combination of both types of asymmetry in populations of many plants, including linden and birch species [6, 7]. The manual or soft measurement of traits only sometimes makes it possible to identify directional asymmetry in the *t*-test. With a normal distribution of the difference (L–R), the DA value is extracted from the difference according to the formula $|L–R| – |L–R|_{av}$, where L and R are the values of the left and right homologous traits, and $|L–R|_{av}$ is the mean sample value L–R by absolute value [8]. Thus, for a researcher who does not set himself testing the features of the leaf shape with an admixture of directional asymmetry, it is quite acceptable to measure linear (angular) traits and determine "pure" FA. The most difficult task is how to determine the shape of the sample distribution. Even with careful collection of leaf blades, each population has a peculiarity in the frequency distribution of values (L–R). What if one of the two samples has a normal Gaussian distribution and the other has a lognormal or exponential distribution? Based on the above task, our goal of the study was to test the "pure" FA of dimensional traits in samples with a deviation from the normal distribution. The task of the work is to compare the FA values between samples with a frequency distribution of the difference (L–R), which differ from the normal distribution, and make appropriate recommendations.

2 Samples and methods

Two populations of the bird cherry (*Prunus padus* L.) were studied in urbanized areas of about 2–3 km^2 in the Moscow region (the city Elektrostal, 55°48′ N, 38°27′ E) and the city of Vladimir (56° 08′00″ N, 40°25′00″ E). For comparison, we used the northern population from the Kola Peninsula (Murmansk oblast, Apatity city, PABSI experimental site, 67°34′ 48′ N, 33°18′ 10′ E).

The Bird cherry leaves of one size were collected in August 2022 from the lower part of tree crowns. The technique collecting, the drying leaves and the photographing is described in a previous work [9]. A three-time measurement in centimeters of only one trait, the half-sheet width, was carried out (by screen digitizer Dig2.31; TPS package, Rholf, 2017). The results were transferred to Excel tables, where primary statistical analyzes were performed, of which we consider the frequency distribution characteristic (L–R) and the normalized difference (FA index) determined by the formula |L–R| / (L+R).

Two-way analysis was carried out using STATISTICA 10 software (Stat. Ink). The section ANOVA - factorial analysis was used. In the same program, distribution fitting was carried out to determine the type of distribution of samples. Nonparametric Kolmogorov-Smirnov analysis (K-S test) was used to determine the normality of the distribution, and Spearman correlation analysis was used to determine the dependence of the FA index on the value of the trait (L+R)/2. A statistical significance level of 95% was applied.

3 Results and discussion

3.1 Descriptive statistics

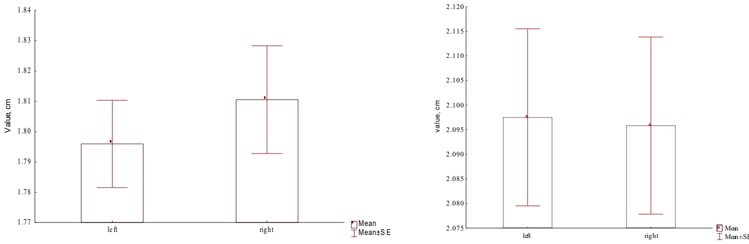
The standard error of measurements of the left and right traits in the Elektrostal population was greater than in the Vladimir population and depended on the magnitude of the variance and the size of the trait itself (larger trait - larger variance - larger standard error). Both distributions (L–R) and FA=|L–R|/(L+R) differed from normal (K-S test: $p<0.05$). These differences are more pronounced in the Elektrostal population. The values of kurtosis and skew of the frequency distribution are presented in Table 1.

Table 1. Descriptive statistics of the samples.

Sample	**n**	**trait, см**	**FA**				
			mean	variance	skew	kurtosis	K-S test *p*<0.05
Vladimir	69	1.81±0.02	0.018±0.004	0.001	4.79±0.29	27.90±0.57	0.33
Electrostal	65	2.10±0.04	0.012±0.002	0.0002	1.92±0.30	3.80±0.59	0.19

The trees in populations did not differ in FA value (F=1.2; p >>0.05), and the value of the trait differs only in the Elektrostal population (df =5; F=15.5; p <<0.01). The correlative statistically significant relationship between the size of the trait and the value of FA has not been evaluated, except for a weakly expressed negative relationship between (L+R) /2 and FA. This made it possible to compare two populations with leaf blades of different sizes. Directional asymmetry was not detected (*t*-test null hypothesis L=R; p >0.05), therefore, such a dimensional trait as the width of half of the leaf plate can be used to determine FA (Fig. 1).

 https://doi.org/10.1051/e3sconf/202342001001



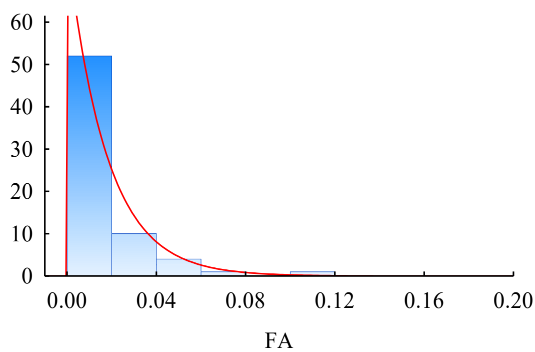
A B

Fig. 1. Mean and standard error (SE) of the left and right trait (A – Vladimir, B – Electrostal).

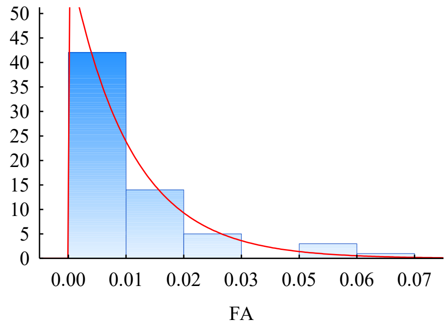
Antisymmetry, as a very rare type of asymmetry, has not been found. We believe that the most appropriate method for determining antisymmetry is to find a correlation between samples L and R. In a nonparametric distribution, Spearman's correlation is usually used. The absence of a significant negative correlation indicates the absence of antisymmetry. In our case, Spearman's *r* coefficient turned out to be statistically insignificant (*r* =0.03; *p* =0.5).

3.2 Frequency distribution and normality

In statistical software, there are over 30 types of frequency distribution can be found. STATISTICA provides seven basic distribution types. We tested samples FA =|L–R| / (L+R), which had a non-parametric distribution. The Kolmogorov-Smirnov test showed a deviation from normality (*p* <0.01-0.02), because the null hypothesis tested the absence of differences between the studied samples and model samples with a normal distribution. The exponential fit of the distribution of FA frequencies with a probability of at least 5% (*p* =0.55 – Elektrostal and *p* =0.13 – Vladimir) was revealed (Fig. 2).



A



B

Fig. 2. Exponential shape of FA frequency distribution (A – Vladimir, B – Electrostal).

Other distributions were not statistically significant. The logarithm with the Box transformation [10] yielded a normal distribution. In this case, the formula d =(|LgL–LgR|+0.00005)$^{0.33}$ was used, where d is the value of the transformed fluctuating asymmetry.

This transformation involves the logarithm of the difference L–R and exponentiation. The FA transformed samples were tested for the normality. The validity of the hypothesis confirmed on over 80%. Checking for a lognormal distribution confirmed the hypothesis over 57%. The *t*-test did not show the mean values differences between the transformed samples (d =0.32 - Vladimir and d=0.33 - Elektrostal). Thus, the distribution of FA after transformation was mostly close to lognormal and normal and indicated a multiplicative error associated with the growth of leaf blades [8]. However, this type of normalization did not allow us to find the desired difference in FA.

3.3 Comparative analysis

The Murmansk population was used for FA comparison. It is characterized by heterogeneity in leaf size (F =9.2; *p* <0.001) and asymmetry at the individual level for individual trees (F =12.9, *p*<0.001). The FA distribution was characterized as exponential (33%) and lognormal (27%), there was no directional asymmetry. The value of FA =|L–R|/(L+R) exceeded the FA of the populations of the middle zone and was equal to 0.024±0.002 (Vladimir – 0.018±0.004; Elektrostal – 0.012±0.002; *p*<0.05). The nonparametric Wilcoxon test determined the difference at *p* =0.001. We also found a negative kurtosis in FA sample (γ = –0.85±0.6), which indicated the small dole of antisymmetry. The heterogeneity in the length of the leaf blades and their individual variability in the northen population, at first glance, increase the asymmetry. As shown by previous studies using the method of geometric morphometrics, the shape of bird cherry leaf blades of the Murmansk population is very variable, and their size is smaller than in the populations of central Russia [2, 9].

The pronounced compensatory leaf blade growth explains a significant standard error of measurement of the left and right halves of bird cherry leaf blades (especially in the Elektrostal population). Associated with this is the high probability of an exponential distribution that normalizes even after simply finding the difference in base 10 logarithms: LgL–LgR. This normalization method is applicable, but not for all samples. If the sample has a high kurtosis, then Box-Cox transformation should be used, which is also not a panacea for non-parametric distribution. In such cases, it is important to avoid outliers that give a "tail" in the distribution of FA values. Thus, the Vladimir population with high values of skewness and kurtosis had a dole of other types of distribution – Poisson or Bernoulli. When previewing herbarium data, the nature of the distribution of L and R sample values is important. High values of variance, skewness, and kurtosis in a FA sample often indicate a deviation from a normal distribution, when comparison with arithmetic means becomes impossible. In a comparative aspect, the dimensional method for determining FA is simple and quite reliable. For the final result, we recommend using the two way ANOVA (individual × side). Its advantage is a high degree of freedom in determining the error, i.e., the proportion of unexplained variance. The results of two-way analysis are shown in Table 2.

Table 2. Result of two way ANOVA.

	SS	***df***	***MS***	***F***	***p***
Vladimir	2.093	68	0.031	3.0	0.000
error	2.811	276	0.010		
Electrostal	0.186	59	0.003	1.1	0.306
error	0.688	240	0.003		

Notes: *SS* – sum of square ; *df* – degree of freedom; *MS* – mean square; *F* –Fisher criterion ; *p* – level of probability

The largest error (FA10; $p > 0.05$) was obtained in the Murmansk population because of the high heterogeneity of the linear parameters of leaf blades. In the Elektrostal population, the large standard error of the sample and the exponentiality of the distribution also led to an increase in the individual × side interaction error. And only one Vladimir population had a statistically significant fluctuating asymmetry. The index FA10 was equal to 0.007 as non-directional asymmetry variance after remowing measurement error is 33% from toal between sides variation. After the Bonferroni correction *p*-level was 0.025.

The pure FA can be obtained, even at high kurtosis values ($\gamma = 2\text{-}3$), but not so high as to cause a large standart error. The deviation from normality, which must be constantly checked, is a consequence of a high kurtosis, i.e. outliers of individual values (variance) outside the confidence interval.

The method of geometric morphometrics confirmed the absence of FA in the Murmansk population; high shape variability here is associated with a low manifestation of asymmetry. This fact does not indicate the stable development of this population, but only characterizes its heterogeneity in the conditions of the northern region.

We recommend for the quantifying testing FA:

1) maximally approximate the size of the leaves collected in different localities;
2) to check the samples (L–R) for the normality;
3) to carry out a triple measurement for determination standard error;
4) to confirm results of the normalizing method in a two-way analysis of variance.

4 Conclusions

a) Traditional morphometry is still useful for determining lamina asymmetry, but can be done using several approaches.
b) If smaller or larger leaves characterize the population, then it is convenient to use the methods of geometric morphometrics.

c) In the northern population, the shape of leaf blades differs significantly (even between two collections), which, however, does not cause variability in developmental stability, estimated by FA.
e) Directional asymmetry is best avoided by a skillful collection of sheet plates. If it is, then you need to check its presence by several methods including 2way ANOVA.

References

1. R.A. Palmer, C. Strobeck, Ann. Rev. Ecol. Syst., **17** (1986)
2. C.P. Klingenberg, Symmetry, **7(2.)** (2015)
3. O.V. Chernakova, G.V. Chudnovskaya, Vestnik ISAA **91** (2019)
4. I.A. Lebedinskiy, C.Y. Mochalova, Vestnik PGGU, **1** (2022)
5. M.A. Suslina, O.S. Kravchenko, L.N. Suntsova, Tech. equip. landscape garden. land. constr (2021)
6. I.E. Zykov, Baranov, E.Yu. Poloskova, I.N. Lipponen, D.D. Kuznetsova, Adv. Current Nat. Sc. Agric. Sc. **7** (2022)
7. S.G. Baranov S.G., T.S. L.I. Bibik, Vestnik KSU, **2** (2017)
8. N.M. Cowart, J. H. Graham Int. J. Plant Sc., **160(1)** (1999)
9. I.E. Zykov, Baranov, E.Yu. Poloskova, I.N. Lipponen, O.A. Goncharova, A.D. Prokopenko, Adv. Current Nat. Sc. Agric. Sc. **4** (2022)
10. J.P. Swaddle, M.S. Witter, I.C. Cuthill, Anim. Behav., **48(4)** (1994)