

A decision matrix to better identify repeatable physiological variation within individuals

Yangfan Zhang^{1,2*}, Chris M. Wood², Colin J. Brauner², Anthony P. Farrell²

¹Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA, USA, 02138

²Department of Zoology, University of British Columbia, 6270 University Boulevard, Vancouver, BC, Canada, V6T 1Z4

Short title: The repeatability of physiological variation in individuals.

Keywords: individual variation, physiological performance, cardiorespiratory system, Bland-Altman analysis, standardized protocols, stable trait, precise measurement

The corresponding author's email: yangfan_zhang@fas.harvard.edu

Summary

The performance of an individual has remained at the heart of evolutionary biology since the time of Darwin. Physiologists are equally drawn to the implications of individual variation for health and sporting endeavours, and specifically, whether or not a physiological trait is repeatable within an individual. Experimental biologists are especially interested in temporally stable physiological traits that are relevant to an individual's lifetime fitness for natural selection to act upon. Experimental noise, however, confounds the measurement of such repeatability, even though validated protocols exist for measuring many meaningful physiological performance traits. Missing is a decision matrix that helps distinguish individual variation from experimental noise. We propose a precision-&-repeatability assessment matrix (PRAM) that integrates established assessments of individual variability and repeatability. This matrix places metrics that are more repeatable and precise in the quadrant closest to the origins of Cartesian coordinates; those farthest away are less acceptable in terms of both repeatability and precision. As a case study, PRAM is applied to whole-organism aerobic and non-aerobic metabolic performance metrics from fish that were measured with the same protocols. The analysis illustrates that aerobic metabolic metrics can be more repeatable and precise than non-aerobic ones. Consequently, PRAM helps physiologists to better understand whether the observed variability is due to non-repeatable metrics or true individual variation.

Glossary

Absolute aerobic scope (AAS): the numerical difference between maximum and standard metabolic rate, defining the aerobic capacity of animals for their activities [1].

Aerobic metabolism: metabolic pathways supported by oxidative phosphorylation under a steady-state condition.

Aerobic metric: a metric measured with a standardized test at the organismic level that estimates either the rate, performance or capacity of aerobic metabolism.

Accumulated oxygen deficit (AOD): Accumulated oxygen deficit is the anaerobic capacity of an animal in oxygen equivalents to be comparable with absolute aerobic scope [2]. It can be measured as the cumulative oxygen deficit when animals are in a severely low oxygen level below critical oxygen saturation and before the animals lose their upright equilibrium.

Anaerobic metabolism: metabolic pathways temporarily supported by substrate-level phosphorylation. Also referred to as non-aerobic at the whole-animal level.

Non-aerobic metric: a metric measured with a standardized test at the organismic level that reflects the performance or capacity of anaerobic pathways. The metric is measured at the organismic level with a standardized test that reflects the performance and capacity based on high-energy phosphate stores, oxygen stores in tissues and substrate-level phosphorylation metabolic pathways.

Bland-Altman analysis: an established statistical method that quantifies measurement variability when a metric is remeasured on the same individual [3].

Critical oxygen saturation: minimum O_2 saturation (critical O_2 saturation, O_{2crit}) needed to sustain standard metabolic rate [4] [5].

Ectotherm: an animal whose body temperature conforms to ambient temperature.

Experimental noise: experimental noise that is attributable to operational, instrumental, methodological, and/or analytical variability. Experimental noise is reduced through better controls (either methodologically or statistically), better testing equipment and better training of personnel who make the measurements.

Excess post-exercise oxygen consumption (EPOC): O_2 requirement to restore tissue O_2 stores and high-energy phosphate stores, and to metabolize the end products of substrate-level phosphorylation after exhaustive exercise [6]. EPOC is another form of accumulated oxygen deficit.

Factorial aerobic scope (FAS): Factorial aerobic scope measures the aerobic capacity as the ratio to the minimum maintenance oxygen demands (maximum O_2 uptake / standard metabolic rate) [7].

Factorial scope for oxygen deficit (FSOD): the ratio between critical oxygen saturation and incipient lethal oxygen saturation [8].

Incipient lethal oxygen saturation (ILOS): oxygen level when the fish loses its upright equilibrium, an assessment of hypoxia tolerance [9].

Individual repeatability: The degree to which a trait in an individual is statistically similar each time it is remeasured. While individual repeatability can be statistically tested, repeatability is typically considered within a specific and defined timescale.

Individual variability: The degree to which a trait numerically varies each time it is remeasured in an individual; The converse of individual repeatability.

Maximum oxygen consumption rate ($\dot{M}O_{2\max}$): the maximum attainable rate of oxygen uptake [6] [10].

Physiological metric: a physiological measurement made at one or multiple levels of a biological system.

Precision: the variability of a repeated measurement; higher numerical variability reflects less precision. Therefore, identifying or improving the precision of measurement improves our confidence in assessing the repeatability of the trait by reducing the risk of Type-II statistical errors so that true biological variation can be detected [11].

Routine metabolic rate (RMR): metabolic rate when animals are in a state of routine activities, which is very much situation-specific [12].

Scope for oxygen deficit (SOD): the numerical difference between critical oxygen saturation and incipient lethal oxygen saturation [8].

Stability: A condition of a metric or trait measured in an individual that exhibits high repeatability. A less variable trait has greater stability. Stability (*see below*) is inferred from a statistical test of a metric's repeatability within an individual.

Standard metabolic rate: the minimum maintenance metabolic rate in the postabsorptive state.

Time instantaneous $\dot{M}O_2$ is above 50% absolute aerobic scope ($T_{>50\%AAS}$): The amount of time an animal spends above 50% of its absolute aerobic scope during measurement of standard metabolic rate [13].

Trait: a well-defined biological metric that has proven mechanistic underpinnings.

Variation: describes the differences in values for a given metric among individuals [14] [15]. Variation is a theoretical concept of frequency distribution pertaining to true biological differences seen among individuals within a population for a metric.

Variability: Variability combines biological variation among a group of individuals and the influence of experimental noise on the values for a given metric, *i.e.*, the full tendency of individual differences in a metric [14] [15] [16].

** The abbreviations listed in the glossary will be directly referred to in the main text.*

Conceptualizing the challenge

Individual variation in physiological performance traits is fundamental for evolution through natural selection [17] [18]. Thus, the performance of individuals ultimately influences the chance of success of a species, a strain, or a population, just as individual human athletes succeed in sports competitions. Consequently, the past century of science has emphasized the importance of individual variation in evolutionary processes [19], and individual-based models have become central in ecology for many decades [20] [21]. Likewise, performance studies in elite athletes have revealed why ‘winners’ are winners, and also provided insight into the upper limits of human physiological performance and the capacity of physiological systems to acclimatize [22]. Beyond human athletes, the physiological limits, capacities and acclimatization potential of other animals are beginning to unfold (*e.g.*, [1] [23–25] [26] [27]). Ultimately, measurement systems that can reliably and precisely estimate the repeatable physiological variation among individuals remain at the heart of both biomedical research and evolutionary biology [17], despite the attraction of "*The Golden Mean*" and a long tradition in scientific culture of minimizing individual variation in our data to obtain a clear picture.

A core challenge for experimental biologists examining the repeatability of individual traits is experimental noise, which exists even when laboratory experiments are highly controlled (Fig. 1). Indeed, if experimental noise is too great when a physiologically meaningful trait is measured and then remeasured for an individual, a Type II statistical error (false negative) can emerge in both longitudinal and cross-sectional studies (because experimental noise equally influences control and treatment groups). Thus, both longitudinal and cross-sectional physiological studies often assume that a trait in an individual is temporally stable over the timescale of the experiment [28].

Yet, physiological systems are dynamic, *e.g.*, ontogeny and aging [29]. Moreover, comparative physiologists who study ectotherms routinely measure this dynamic when measuring acclimation or acclimatization responses to new environmental conditions. Therefore, while comparative physiologists have necessarily examined the temporal stability of a physiological trait in “control” animals, the understanding of the repeatability of physiological traits under “experimental” conditions (where the animal is challenged to perform) lags well behind that of sports and biomedical physiology [*e.g.*, exercise performance [30] [31], temperature regulation [32] [33] [34] [35], and overall metabolism [36] [37] [38] [39].

Consequently, the major goal of this study is to propose and test an analytical approach that would help comparative physiologists better separate those traits within an individual that are repeatable, from those traits confounded by too much experimental noise. Herein, we propose a Precision-&-Repeatability Assessment Matrix (PRAM) that integrates established statistical methods that can measure variability and individual repeatability (*see* Glossary for important distinction). Furthermore, our case study, which focuses on whole-organism physiological performance, is a suitable candidate for all animals because of its established relationship to locomotion performance [40], its important implications on lifetime fitness [41] [42], and its demonstrated genetic underpinnings in some ectothermic vertebrate systems [27] [43].

In particular, metabolic rate, typically estimated by measuring oxygen (O₂) uptake under various conditions (*e.g.*, basal, routine, or maximum sustainable O₂ uptake rates) and environments, has been broadly used to assess the robustness of the vertebrate cardiorespiratory system [44] [45] and even linked to human longevity [46] [47] [48]. We also consider metrics of whole-animal anaerobic metabolism, the glycolytic pathways that utilize substrate-level phosphorylation to generate ATP, which in contrast, play a crucial role in shorter-term, life-promoting situations [6] [2] [49] [50] [51] [52] [53], such as environmental hypoxia [54] [2] [55], predation and predator-escape responses [56] [57] [58] [59], and difficult reproductive migration passages [60] [23]. Indeed, anaerobic metabolic metrics have received scant attention; just 12% of published repeatability studies in sports medicine and comparative physiology (summarized in Table S1) have examined indirect metrics to make largely inferential assessments of whole-animal anaerobic performance (Table S1). Here, we take advantage of the standardized Integrated Respiratory Assessment Protocol (IRAP, *see* supplementary materials for details) that measures both aerobic and anaerobic metabolic metrics on individual fish. No single study to date has compared the repeatability of aerobic and non-aerobic metabolic metrics. Therefore, our secondary goal is to compare the repeatability of aerobic and non-aerobic metrics for whole-animal metabolism measured by IRAP.

The focus of comparative physiology on the metabolic rate of fishes

Comparative physiologists study perhaps the most diverse biological system among all experimental biologists, which in turn is well poised to tackle the challenges of studying biological variations. Among all the physiological performance traits studied in vertebrates, a great deal is known about the precision and reliability of whole-animal metabolic

measurement techniques in fish [1] [6] [61] [62] [63] [4] [64] [65] [66] [67] [68] [10] [69] [70] [71] [72]. Moreover, fish species are the most diverse vertebrate lineage (over 30,000 species), representing over 50% of vertebrate species by number and inhabiting almost every aquatic environment on the planet [73] [74]. Their physiological systems, as a result, are replete with biological variation and acclimation potential. Even a fundamental physiological trait such as blood O₂ transport can vary enormously among fish species [75] [76], which have the only known vertebrate lineage without haemoglobin (*e.g.*, icefish [77]). Within an individual, different haemoglobin isoforms can change over periods of three weeks [78], and factors that alter the oxygen affinity of hemoglobin can be modified within hours [79] [80].

Yet, comparative physiologists have continued to package individual variation around a mean value bounded by the standard error bars rather than exploring the potential significance of this individual variation, namely the ‘Tyranny of the ‘Golden Mean’ [17]. Although the reasons for the emphasis on the ‘Golden Mean’ in comparative physiology remain valid and multifaceted [81], gains are being made in acknowledging individual variation, reporting the precision of measurements [11] and reporting experimental conditions to better assess the methodologies [81]. (Table 1 summarizes the challenges of pursuing high-quality experimental studies of physiological performance alongside immediate and long-term solutions.). Therefore, we propose that comparative physiology and fish physiology specifically can benefit from the application of PRAM, which incorporates some of the key operational principles used in biomedical physiology and animal ecology.

Approaches used by other disciplines to account for individual variation

Biomedical physiologists typically identify a healthy human state within a range of values for a given parameter that is normal for individuals (rather than a single mean of a population), an approach not systematically explored by comparative physiologists [82] [11]. The presumption with this approach is that the parameter can be measured precisely.

A Bland-Altman analysis, a statistical approach to assess individual variation [3] [83], is often used in sports physiology. It calculates the percentage difference between two measurements taken from the same individual, divided by their average [3]. The percentage difference measured on the same individual and the average value become a single data point on a Bland-Altman plot (Fig. 2). Experimental noise, however, is incorporated into a Bland-Altman analysis of individual variation.

The normalized coefficient of repeatability (COR) for a group of individuals quantifies the variability of a metric (Fig. 2) (Eqn. 1) [83].

Eqn. 1:
$$\text{Normalized COR} = \frac{2 \times \sqrt{[\sum(\chi_i - \mu)^2]/(N-1)}}{(\alpha + \beta)/2} \times 100\%$$

where χ_i is each value of the difference between two repeated measurements from each individual, μ is the mean of the difference between the two repeated measurements, N is the sample size, α is the value of a metric measured in the 1st test, and β is the value of a metric measured in the 2nd test.

Repeatability of rank order in a repeated measurement design has been used by comparative physiologists to examine if the individuals maintain their relative rank within a population despite a shift in the population average (Also see the rank order analyses of our dataset in Figs. S1, S2) [84] [85] [86] [87] [88] [89]). Analysis of rank order, however, loses the quantitative aspect of the metric under consideration: the winner can remain the winner, but information on the absolute performance is lost.

Generalized Linear Mixed Models (GLMM) are often used by animal ecologists and ecophysiologists working in field settings who face a greater challenge than laboratory-based experimental biologists concerning the ability/inability to control important environmental variables, e.g., temperature, rainfall, humidity, salinity, oxygen levels, and photoperiod. Uncontrolled environmental variables add to experimental measurement “noise”. In this case, a Generalized Linear Mixed Model (GLMM) statistical approach can account for the impacts of variables that cannot be controlled [90] [91] [92] and the contribution of the experimental noise [93] [94] when testing the repeatability of measured metrics. A GLMM, however, does not attempt to separate experimental measurement noise from true individual variation.

Precision-&-Repeatability Assessment Matrix (PRAM)

In view of the limitations for each of the statistical approaches noted above (Table S2), we propose a Precision-&-Repeatability Assessment Matrix (PRAM) that graphically combines a parametric test of repeatability (the Pearson correlation coefficient) with a normalized COR. This theoretical framework then provides a quantitative approach for experimental biologists to better visualize the repeatability and variability of metrics (Fig. 3).

PRAM statistically quantifies variability within the individual-based measurements using the well-established Bland-Altman analysis [95] [3] [96] [30]. Also, correlation analysis (e.g. Pearson’s correlation coefficients) assesses the repeatability of a given metric measured twice on the same individual using a balanced design [97] [90]. These two statistical analyses are then placed on Cartesian coordinates to arrive at a decision matrix with four quadrants (Fig. 3) that inform the quality of the performance metric by parsing out the

biological variation among a group of individuals and experimental noise. For example, the quadrant closest to the origin identifies a metric that is stable and measured by precise protocols (*i.e.*, resulting in low experimental noise). These repeatable metrics would be best suited for testing hypotheses, developing theories, and building models with strong predictive power. Moving away from the origin of the quadrant, the other quadrants represent metrics that are less repeatable for different reasons, *i.e.*, precise measurements applied to unstable traits, imprecise measurements applied to stable traits, and imprecise measurements applied to unstable traits (Fig. 3). Therefore, the decision matrix of PRAM helps distinguish experimental noise from individual variation, enabling a better understanding of physiological variation.

A case study: applying PRAM to whole-animal metabolic metrics for fish

The data used for the case study were those collected from two independent repeated-measure design studies with two genetically distant athletic fish species, Atlantic salmon (*Salmo salar*) and European sea bass (*Dicentrarchus labrax*) [98] [99] [100]. Both studies retested the same individuals after ~4 weeks (*see* Supplementary Materials for animal information). Moreover, both used IRAP, which holistically characterizes an individual fish's respiratory phenotype by measuring up to 13 metrics for the performance and capacity of aerobic and anaerobic metabolism [8] (*see* Glossary for full definitions of terms & supplementary material for details of their measurement, Fig. S3, S4). Consequently, the same practitioners performed the standardized IRAP procedure under controlled laboratory conditions and with similar equipment (Table S3), which minimized operator variability (*see* Table 1). Also, phenotype stability assessment was conducted over a 1-2 month time frame, acknowledging that fish can partially or fully acclimate to new environmental conditions over several weeks [101] [102] [103] and that only similar life stages should be compared in the first instance.

In principle, the temporally stable traits measured precisely should be repeatable across species. Otherwise, the targeted trait either does not have a consistent stability among individuals within in species, or the measurement is imprecise. PRAM was then applied to identify which individual traits were stable, which were not and which were noisy (Fig. 4).

When COR was less than 53%, a metric was nearly three times likely to be repeatable (the peak of the frequency distribution of arbitrary p -value < 0.05 was nearly triple the peak of the frequency distribution of p -value ≥ 0.05 , Fig. 4a). When COR was greater than 53%, the chance of a metric being unrepeatable increased (Fig. 4a). Therefore, we used COR of

53% as a nominal line of reference for the measurement precision in this analysis (Note: 53% is applicable for this dataset and a work in progress). The working criteria to set the threshold value are 1) the transition point in the frequency distribution of COR where metrics start to shift from a higher chance (e.g. 2-3 times more likely) being repeatable to a lower (e.g. 0.5 times) chance being repeatable. We then set the quadrants for PRAM arbitrarily as COR of 53% for measurement precision (green vertical dashed line on the x-axis) and repeatability (green horizontal dashed line on the y-axis) (Fig. 4b). As perhaps expected, the more precisely measured metrics (*i.e.* smaller COR values) tended to be more repeatable (correlation test: $p\text{-value} \leq 0.01$, Fig. 4b), but with some exceptions, *e.g.*, FSOD was measured precisely and was not repeatable (Fig. 4b).

PRAM also demonstrated that individual aerobic metabolism traits (AAS, FAS, $\dot{M}O_{2\max}$, SMR, and RMR) were repeatable for both species of athletic fish. The only unrepeatable aerobic metric was $T_{>50\%AAS}$, an index of spontaneous activity that is likely context-dependent. $\dot{M}O_{2\max}$ is an important independent metric to derive AAS & FAS. Bland-Altman analysis on $\dot{M}O_{2\max}$ demonstrated that $\dot{M}O_{2\max}$ has a limited measurement variability (the narrow horizontal band in Fig. 2; Bland-Altman analyses of other metrics are in Fig. S5 & S6), a finding that may reflect past efforts to establish reliable measurement techniques for $\dot{M}O_{2\max}$ [10] [69] [104] [45] and the tightly coupled steps of the O_2 transport cascade for peak performance in vertebrates [105] [106]. When measured properly, $\dot{M}O_{2\max}$ should have limited intra-individual variation in a repeated measure design, as previously shown in healthy humans [83] [107] [106] [22]. Thus, PRAM analysis confirmed that measurement precision and trait stability were necessary features of the well-established independent performance metrics. Indeed, studies with other animal taxa find strong repeatability of aerobic metabolic metrics (Table S1). Also, SMR, $\dot{M}O_{2\max}$, AAS and FAS have been related to the lifetime fitness of vertebrate species in various environments [108] [109] [110] [111] [27], with some but not all [112] studies finding stable traits within a life stage.

A novel discovery of the PRAM analysis was that whole-organism anaerobic metabolism metrics were generally either unstable or were imprecisely measured. For example, EPOC, a metric that predominantly measures the oxygen equivalents of glycolytic metabolism during exhaustive exercise, was reported as both a stable and an unstable metric. This discovery may reflect the many challenges that still exist to properly quantify anaerobic metabolism at the organismal level (*see* reviews [113] [114] [115] [116]), those beyond the extent to which glycolysis is affected by individual variation in cardiorespiratory robustness [71] [72] [2] [117] [100]. For example, differences in phosphagen and metabolic substrate

stores, as well as glycogenesis, can all potentially introduce variation [118] [119] [120]. Similar concerns also apply to the scope for oxygen deficit, a new performance proxy for the anaerobic capacity of a fish in a hypoxia challenge test, [8] which quantifies the accumulation of glycolytic end-products and different demands on glycolysis in severe hypoxia [8] [121]. The Bland-Altman analysis for the scope for oxygen deficit revealed large variability (a wider horizontal band in Fig. 2; see Bland-Altman analyses for other metrics in Fig. S5 & S6). Thus, assessment protocols for whole-animal anaerobic capacity and performance, which remain in their infancy, perhaps should be given greater attention to develop more precise methods, especially given that they fuel shorter-term, life-saving activities.

Surprisingly, PRAM analysis revealed important nuances for two widely used and well-established metrics used to estimate hypoxia tolerance in fish, *i.e.*, O_{2crit} and ILOS, which were measured with the same standardized hypoxia challenge test [9]. While O_{2crit} was precisely measured (*i.e.* low experimental noise) in both species, ILOS was precise only in Atlantic salmon (*Salmo salar*) and not in European sea bass (*Dicentrarchus labrax*) (Fig. 4b). O_{2crit} , as calculated here, benefits from an objective mathematical approach (SMR is the baseline to derive an intercept with a linear regression equation [64] [5]). In contrast, ILOS is a visually subjective recording of the ambient O_2 level when a fish loses its upright equilibrium. Why O_{2crit} demonstrated inter-specific variation [67] for its repeatability between *S. salar* but not *D. labrax* is unclear.

Future directions for research on the repeatability of individual variation in experimental biology

Our case study of PRAM using IRAP metrics for two species of ray-finned fish is only a demonstration that informs us of a few key general principles for experimental biology. An unstable trait can be unrepeatably even when precisely measured (Quadrant II in PRAM; Fig. 3). Conversely, and as expected, a stable trait requires a standardized protocol to generate precise measurements (*i.e.*, low experimental noise) (Quadrant III in PRAM; Fig. 3). Being able to distinguish the two scenarios is important for biological experiments, and this is what PRAM achieved. The same principles likely apply to other performance metrics in other animal and human models, and perhaps across multiple biological organizations.

Striving for repeatability has been a common goal among experimental biologists in recent literature [122] [123] [124] [125] [11]. The central idea behind the PRAM is to characterize the repeatability and measurement variability of a metric on a decision matrix to distinguish the true physiological variation from experimental noise. PRAM assesses the

quality of the metrics and informs the types of conclusions that can be drawn. Much work remains, however, to identify the multifaceted physiological reasons for the differences in repeatability for metabolic metrics, as well as further analyses of the PRAM framework in acclimation (or acclimatization) studies. For example, how certain acclimation states can lead to less stable traits (e.g. transition from quadrant I to quadrant II). Hence, we welcome future improvements to the decision matrix, given that statistical assessment methods are constantly evolving. For example, the quantification of variance can use the coefficient of variation or covariance [126] [91], and the assessment of repeatability can use residual variance [90]. For repeatability tests conducted on more than two time points (e.g. [127] [58] also see reviews on this topic: [126] [128]), the decision matrix can identify the sources of the variability in a pair-wise manner across multiple time sampling points. The criteria for the COR threshold that determine a precise measurement are a work in progress. It remains to be tested with future studies whether a more rigid COR threshold can be established (e.g., ~50% COR).

The PRAM framework can also contribute to two long-standing questions in experimental biology: 1) How to mitigate the trade-offs between sample size and measurement precision to improve statistical power at a limited experimental resource? 2) How to account for the learning behaviours over multiple repeatability tests? First, if the imprecise measurement (quadrant III) quantifies the stable traits, a large sample size can potentially compensate for the statistical power. Still, increasing the sample size of imprecise measurement would be less productive than trying to source and reduce the noise. If the measurements are precise and targeting stable traits (quadrant I), a smaller sample size might be needed to reach the desired statistical power. The precise metrics would better utilize the limited research resources when quantifying the plastic traits (e.g., quadrant II). Moreover, our PRAM framework could be used to retrospectively analyze existing data, as shown here, to identify metrics that appear robust (quadrant I) and in need of some scrutiny (quadrants III & IV), as well as either rejecting those metrics in quadrant IV or investigating the underlying issues. One potential issue worth investigating is the importance of learning behaviour in the quality of a metric. Using the PRAM framework as part of an experimental design can help answer the question of how many repeatability tests and over what time intervals are necessary.

Overall, PRAM analysis will help comparative physiologists and experimental biologists to objectively, quantitatively, visually and more systematically understand the metrics they measure. PRAM can be applied well beyond the limited number of metrics used here as a case study. Furthermore, PRAM can better calibrate the confidence placed in the

386 indicators that we believe are metrics of fundamental metabolic traits. This analytical
 387 approach can enable physiological metrics to be better integrated into studies of genomics,
 388 molecular mechanisms, biomechanics, the life history of the organisms, and the
 389 environmental selection pressure over the evolutionary history. After all, inter-individual
 390 variation is the raw material for evolution through natural selection.

Ethics: This work has ethical approvals from animal welfare committees (*see supplementary methods for the certificate numbers of the animal welfare approvals*).

Data accessibility: Data are archived in a [repository](#) for publication.

Funding information: Y.Z. is supported by a Postdoctoral Fellowship of the Natural Sciences and Engineering Research Council of Canada (NSERC PDF-557785–2021), followed by a Banting Postdoctoral Fellowship (202309BPF-510048-BNE-295921) of NSERC & CIHR (Canadian Institutes of Health Research). C.M.W.’s research is supported by an NSERC Discovery Grant (RGPIN-2023-03714). C.J.B.’s research is supported by an NSERC Discovery Grant (RGPIN-2023-03456). A.P.F. was funded by NSERC and a Canada Research Chair.

Acknowledgments: Many thanks to members of the Department of Zoology at the University of British Columbia and George Lauder at Harvard University for numerous discussions about individual variation and the repeatability of metrics. To Dave Randall’s question of what defines a trait.

Author contributions: Y.Z., C.M.W., C.J.B., A.P.F. conceptualized the study. Y.Z. performed experiments and data analyses, and wrote the original manuscript. All authors provided manuscript edits and comments and approved the final version.

Competing interests: The Authors declare that they have no competing interests.

410 **Table 1. Challenges and solutions to improve data quality.** Three general areas of challenge that collectively contribute to
 411 experimental noise are instruments, biology and operators. For each area, suggestions for both short- and long-term solutions are
 412 offered.

413

Challenge	Possible immediate solution	Possible long-term solution
Instrument error	Standardize and improve experimental instrumentation	Funding of university-industry collaborations to improve instruments
Measurement error vs	• Less stressful and invasive measurement techniques	More funding and less restrictions for exploratory research
Biological variation	• Standardize approaches for data analysis • Standardize protocols	
Operator error	Improve individual training	Streamline training

414

Fig. 1. Schematic of the effects of methodological errors on measuring stable traits.

When measuring a stable trait, the protocol, electronic, and analytical errors can inflate the variability of the measurement (for simplicity, the experimental noise is assumed to be equally distributed around the true mean in the schematic) and mask the real biological variation. At a new steady state, the distribution of biological variation can shift to a different level. If instrumental, protocol, and analytical errors are sufficient, measurement variability can overlap amply to compromise the precision in quantifying the real biological variation. The principal idea is to eliminate sources of error as best as possible to detect 1) what the real biological variation is and 2) any shifts in the true value of an individual metric.

Fig. 2. Bland-Altman analyses of repeatable and unrepeatable metabolic metrics characterized by the standardized Integrated Respiratory Assessment Protocol (IRAP), which measures whole-animal aerobic and non-aerobic metabolic metrics. Metrics were remeasured 4 weeks apart on the same individuals of European sea bass (*Dicentrarchus labrax*) at 25 °C (~30 ppt, ~60 g, n=16) and Atlantic salmon (*Salmo salar*) at 11 °C (~30 ppt, ~75 g, n=16). Metrics in blue were repeatable; metrics in red were unrepeatable (as determined by the Pearson correlation coefficient test using $p < 0.05$ (r -value: p -values appeared in the upper right corner of each figure). Coloured shading indicates 95% limits of agreement, a measure of the metric variability (twice its standard deviation and accounting for the mean difference of the two measurements). Here, as representatives, an aerobic metabolic metric is the maximum oxygen uptake rate ($\dot{M}O_{2max}$) and a non-aerobic metabolic metric is scope for oxygen deficit (SOD). The repeatable aerobic metabolic metric has a much narrower band of intra-individual variability (a smaller value of 95% limits of agreement) than the unrepeatable non-aerobic metabolic metric. The data were adapted from the published studies [98] [100].

Fig. 3. Theoretical framework for measurement precision and trait stability using a Precision-&-Repeatability Assessment Matrix (PRAM). The decision matrix is an analytical concept and approach based on the association between the variability of metrics and the repeatability of the metrics. The utility of the decision matrix is to distinguish the stability of traits in individuals from the experimental noise. The normalized coefficient of repeatability (COR) is a proxy for measurement precision at the x-axis, and the p -value from correlation coefficient tests is a proxy for repeatability at the y-axis. This forms quadrants of interest: i) precise measurements and stable traits; ii) precise measurements and unstable traits; iii) imprecise measurements and stable traits; iv) imprecise measurements and unstable traits. The area of transition for repeatability is a p -value of 0.05 using correlation coefficient tests. The area of transition for precision is the normalized coefficient of repeatability value with threshold values corresponding to a higher probability for the metrics to become repeatable (e.g. p -value < 0.05).

Fig. 4. An examination of the relationship between the variabilities of metabolic metrics with the repeatability of the metabolic metrics. (a) The frequency distributions of the p -value of Pearson correlation coefficient tests (categorized into groups of Pearson $p < 0.05$ and Pearson $p \geq 0.05$) informed the formation of quadrants in the Precision-&-Repeatability Assessment Matrix (PRAM; see Fig. 3). The green vertical dashed lines [normalized coefficient of repeatability (COR) = 53%] are based on the frequency distribution of the p -value of Pearson correlation coefficient tests. When the metrics have COR values of less than 53% (between the origin and the green vertical dashed lines), the metrics have nearly three times higher probabilities of being repeatable ($p < 0.05$). When COR is larger than 53% (beyond the green vertical dashed lines), the metrics had a lower chance of being repeatable. (b) The green vertical dashed lines annotate the threshold (COR = 53%) where metrics

change from being precise to imprecise. The p -value of 0.05 is annotated by the green horizontal dashed lines to indicate the area where the metrics gradually become unrepeatable from being repeatable. Using these arbitrary threshold values of repeatability and precision, quadrants were formed in panels using the theoretical framework of PRAM: i) precise measurements and stable traits; ii) precise measurements and unstable traits; iii) imprecise measurements and stable traits; iv) imprecise measurements and unstable traits. Different symbols denote different metabolic metrics. To understand the general relationship between variabilities and repeatability, the analyses combined the measurements obtained from European sea bass (*Dicentrarchus labrax*) at 25 °C (~30 ppt, ~60g, n=16; filled symbols) and Atlantic salmon (*Salmo salar*) at 11 °C (~30 ppt, ~75g, n=16; half-filled symbols). Metrics in blue measure traits when animals predominantly rely on aerobic metabolism (*i.e.* aerobic traits), and metrics in red measure traits when animals engage in a substantial amount of glycolysis (*i.e.* non-aerobic traits). Aerobic metabolic metrics (in blue) include absolute aerobic scope (AAS), factorial aerobic scope (FAS), maximum oxygen uptake ($\dot{M}O_{2max}$), standard metabolic rate (SMR), routine metabolic rate (RMR), and time spent above 50% AAS ($T_{> 50\% AAS}$). Non-aerobic metabolic metrics (in red) include excess post-exercise oxygen consumption (EPOC), scope for oxygen deficit (SOD), the factorial scope for oxygen deficit (FSOD), incipient lethal oxygen saturation (ILOS), accumulated oxygen deficit (AOD) and critical oxygen saturation (O_{2crit}). In general, less variability in measuring the metrics is positively related to the higher probabilities of being repeatable metrics. The data were adapted from the published studies [98] [100].

References

1. Fry FEJ. 1947 *Effects of the environment on animal activity*. Toronto: Univof Toronto Press.
2. Hochachka PW. 1990 Scope for survival: a conceptual “mirror” to fry’s scope for activity. *Transactions of the American Fisheries Society* **119**, 622–628. (doi:10.1577/1548-8659(1990)119<0622:SFSACT>2.3.CO;2)
3. Bland JM, Altman DG. 1986 Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**, 307–310.
4. Fry FEJ, Hart JS. 1948 The relation of temperature to oxygen consumption in the goldfish. *The Biological Bulletin* **94**, 66–77. (doi:10.2307/1538211)
5. Claireaux G, Chabot D. 2016 Responses by fishes to environmental hypoxia: integration through Fry’s concept of aerobic metabolic scope. *Journal of Fish Biology* **88**, 232–251. (doi:10.1111/jfb.12833)
6. Brett JR. 1964 The respiratory metabolism and swimming performance of young sockeye salmon. *J. Fish. Res. Bd. Can.* **21**, 1183–1226. (doi:10.1139/f64-103)
7. Clark TD, Ryan T, Ingram BA, Woakes AJ, Butler PJ, Frappell PB. 2005 Factorial aerobic scope is independent of temperature and primarily modulated by heart rate in exercising murray cod (*Maccullochella peelii peelii*). *Physiological and Biochemical Zoology: Ecological and Evolutionary Approaches* **78**, 347–355. (doi:10.1086/430034)
8. Zhang Y, Mauduit F, Farrell AP, Chabot D, Ollivier H, Rio-Cabello A, Le Floch S, Claireaux G. 2017 Exposure of European sea bass (*Dicentrarchus labrax*) to chemically dispersed oil has a chronic residual effect on hypoxia tolerance but not aerobic scope. *Aquatic Toxicology* **191**, 95–104. (doi:10.1016/j.aquatox.2017.07.020)
9. Claireaux G, Théron M, Prineau M, Dussauze M, Merlin F-X, Le Floch S. 2013 Effects of oil exposure and dispersant use upon environmental adaptation performance and fitness in the European sea bass, *Dicentrarchus labrax*. *Aquatic Toxicology* **130–131**, 160–170. (doi:10.1016/j.aquatox.2013.01.004)
10. Zhang Y, Gilbert MJH, Farrell AP. 2019 Finding the peak of dynamic oxygen uptake during fatiguing exercise in fish. *Journal of Experimental Biology* **222**, jeb196568. (doi:10.1242/jeb.196568)
11. Williams S, Carson R, Tóth K. 2023 Moving beyond P values in The Journal of Physiology: A primer on the value of effect sizes and confidence intervals. *The Journal of Physiology* **601**, 5131–5133. (doi:10.1113/JP285575)
12. Wagner GN, Kuchel LJ, Lotto A, Patterson DA, Shrimpton JM, Hinch SG, Farrell AP. 2006 routine and active metabolic rates of migrating adult wild sockeye salmon (*Oncorhynchus nerka* Walbaum) in Seawater and Freshwater. *Physiological and Biochemical Zoology: Ecological and Evolutionary Approaches* **79**, 100–108. (doi:10.1086/498186)
13. Zhang Y, Polinski MP, Morrison PR, Brauner CJ, Farrell AP, Garver KA. 2019 High-load reovirus infections do not imply physiological impairment in salmon. *Front. Physiol.* **0**. (doi:10.3389/fphys.2019.00114)

- 523 14. Wagner GP, Altenberg L. 1996 Perspective: Complex adaptations and the evolution of
524 evolvability. *Evolution* **50**, 967–976. (doi:10.1111/j.1558-5646.1996.tb02339.x)
- 525 15. Wagner GP, Booth G, Bagheri-Chaichian H. 1997 A population genetic theory of canalization.
526 *Evolution* **51**, 329–347. (doi:10.1111/j.1558-5646.1997.tb02420.x)
- 527 16. Burton T, Killen SS, Armstrong JD, Metcalfe NB. 2011 What causes intraspecific variation in
528 resting metabolic rate and what are its ecological consequences? *Proceedings of the Royal*
529 *Society B: Biological Sciences* **278**, 3465–3473. (doi:10.1098/rspb.2011.1778)
- 530 17. Bennett AF. 1987 Interindividual variability: an underutilized resource. *New Directions in*
531 *Ecological Physiology*. , 147–169.
- 532 18. Feder ME, Bennett AF, Huey RB. 2000 Evolutionary Physiology. *Annual Review of Ecology and*
533 *Systematics* **31**, 315–341. (doi:10.1146/annurev.ecolsys.31.1.315)
- 534 19. Darwin C. 1872 The origin of species by means of natural selection: or, the preservation of
535 favoured races in the struggle for life. D. Appleton and Company.
- 536 20. Botkin DB, Janak JF, Wallis JR. 1972 Some ecological consequences of a computer model of
537 forest growth. *Journal of Ecology* **60**, 849–872. (doi:10.2307/2258570)
- 538 21. DeAngelis DL, Grimm V. 2014 Individual-based models in ecology after four decades.
539 *F1000Prime Rep* **6**, 39. (doi:10.12703/P6-39)
- 540 22. Schumacher YO, Vogt S, Roecker K, Schmid A, Coyle EF. 2005 Scientific considerations for
541 physiological evaluations of elite athletes. *Journal of Applied Physiology* **99**, 1630–1631; author
542 reply 1631-1632. (doi:10.1152/japplphysiol.00563.2005)
- 543 23. Pörtner HO, Farrell AP. 2008 Physiology and climate change. *Science* **322**, 690–692.
- 544 24. Korsmeyer KE, Dewar H, Lai NC, Graham JB. 1996 The aerobic capacity of tunas: Adaptation for
545 multiple metabolic demands. *Comparative Biochemistry and Physiology Part A: Physiology* **113**,
546 17–24. (doi:10.1016/0300-9629(95)02061-6)
- 547 25. Gebczyński AK, Konarzewski M. 2009 Metabolic correlates of selection on aerobic capacity in
548 laboratory mice: a test of the model for the evolution of endothermy. *J Exp Biol* **212**, 2872–
549 2878. (doi:10.1242/jeb.030874)
- 550 26. Schulte PM. 2015 The effects of temperature on aerobic metabolism: towards a mechanistic
551 understanding of the responses of ectotherms to a changing environment. *Journal of*
552 *Experimental Biology* **218**, 1856–1866. (doi:10.1242/jeb.118851)
- 553 27. Prokkola JM, Åsheim ER, Morozov S, Bangura P, Erkinaro J, Ruokolainen A, Primmer CR,
554 Aykanat T. 2022 Genetic coupling of life-history and aerobic performance in Atlantic salmon.
555 *Proceedings of the Royal Society B: Biological Sciences* **289**, 20212500.
556 (doi:10.1098/rspb.2021.2500)
- 557 28. Zhang Y, Mauduit F, Pettinau L, Ollivier H, Lancien F, Anttila K, Farrell AP, Claireaux G. 2023
558 Respiratory plasticity during acclimation to hypoxia and following a recovery in normoxia. *Can.*
559 *J. Zool.* **101**, 794–806. (doi:10.1139/cjz-2022-0158)

29. Zhang Y, Anttila K, Hickey AJ. 2024 Editorial: Exploring the developmental plasticity and transgenerational effects on the thermal biology of aquatic ectotherms. *Front. Physiol.* **15**. (doi:10.3389/fphys.2024.1507027)
30. Hopkins WG. 2000 Measures of reliability in sports medicine and science. *Sports Med* **30**, 1–15. (doi:10.2165/00007256-200030010-00001)
31. Lolli L, Batterham AM, Weston KL, Atkinson G. 2017 Size exponents for scaling maximal oxygen uptake in over 6500 humans: A systematic review and meta-analysis. *Sports Med* **47**, 1405–1419. (doi:10.1007/s40279-016-0655-1)
32. Stevenson RD. 1985 Body size and limits to the daily range of body temperature in terrestrial ectotherms. *The American Naturalist* **125**, 102–117.
33. Mee JA, Doust J, Maxwell NS. 2015 Repeatability of a running heat tolerance test. *Journal of Thermal Biology* **49–50**, 91–97. (doi:10.1016/j.jtherbio.2015.02.010)
34. Zaproudina N, Varmavuo V, Airaksinen O, Närhi M. 2008 Reproducibility of infrared thermography measurements in healthy individuals. *Physiol. Meas.* **29**, 515. (doi:10.1088/0967-3334/29/4/007)
35. Artacho P, Jouanneau I, Le Galliard J-F. 2013 Interindividual variation in thermal sensitivity of maximal sprint speed, thermal behavior, and resting metabolic rate in a lizard. *Physiol Biochem Zool* **86**, 458–469. (doi:10.1086/671376)
36. Speakman JR. 1998 The history and theory of the doubly labeled water technique¹². *The American Journal of Clinical Nutrition* **68**, 932S–938S. (doi:10.1093/ajcn/68.4.932S)
37. Rumpler W, Seale J, Conway J, Moe P. 1990 Repeatability of 24-h energy expenditure measurements in humans by indirect calorimetry. *The American Journal of Clinical Nutrition* **51**, 147–152. (doi:10.1093/ajcn/51.2.147)
38. Wong WW, Roberts SB, Racette SB, Das SK, Redman LM, Rochon J, Bhapkar MV, Clarke LL, Kraus WE. 2014 The doubly labeled water method produces highly reproducible longitudinal results in nutrition studies. *The Journal of Nutrition* **144**, 777–783. (doi:10.3945/jn.113.187823)
39. Trabulsi J, Troiano RP, Subar AF, Sharbaugh C, Kipnis V, Schatzkin A, Schoeller DA. 2003 Precision of the doubly labeled water method in a large-scale application: evaluation of a streamlined-dosing protocol in the Observing Protein and Energy Nutrition (OPEN) study. *Eur J Clin Nutr* **57**, 1370–1377. (doi:10.1038/sj.ejcn.1601698)
40. Garland T, Else PL. 1987 Seasonal, sexual, and individual variation in endurance and activity metabolism in lizards. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **252**, R439–R449. (doi:10.1152/ajpregu.1987.252.3.R439)
41. Biro PA, Stamps JA. 2010 Do consistent individual differences in metabolic rate promote consistent individual differences in behavior? *Trends in Ecology & Evolution* **25**, 653–659. (doi:10.1016/j.tree.2010.08.003)
42. Biro PA, Garland T, Beckmann C, Ujvari B, Thomas F, Post JR. 2018 Metabolic scope as a proximate constraint on individual behavioral variation: Effects on personality, plasticity, and predictability. *Am Nat* **192**, 142–154. (doi:10.1086/697963)

- 599 43. Prokkola JM, Chew KK, Anttila K, Maamela KS, Yildiz A, Åsheim ER, Primmer CR, Aykanat T.
600 2024 Tissue-specific metabolic enzyme levels covary with whole-animal metabolic rates and
601 life-history loci via epistatic effects. *Philosophical Transactions of the Royal Society B: Biological*
602 *Sciences* **379**, 20220482. (doi:10.1098/rstb.2022.0482)
- 603 44. Baba R, Tsuyuki K, Kimura Y, Ninomiya K, Aihara M, Ebine K, Tauchi N, Nishibata K, Nagashima
604 M. 1999 Oxygen uptake efficiency slope as a useful measure of cardiorespiratory functional
605 reserve in adult cardiac patients. *Eur J Appl Physiol* **80**, 397–401. (doi:10.1007/s004210050610)
- 606 45. Hawkins MN, Raven PB, Snell PG, Stray-Gundersen J, Levine BD. 2007 Maximal oxygen uptake
607 as a parametric measure of cardiorespiratory capacity. *Med Sci Sports Exerc* **39**, 103–107.
608 (doi:10.1249/01.mss.0000241641.75101.64)
- 609 46. Strasser B, Burtcher M. 2018 Survival of the fittest: VO₂max, a key predictor of longevity?
610 *Front Biosci (Landmark Ed)* **23**, 1505–1516. (doi:10.2741/4657)
- 611 47. Joyner MJ, Green DJ. 2009 Exercise protects the cardiovascular system: effects beyond
612 traditional risk factors. *The Journal of Physiology* **587**, 5551–5558.
613 (doi:10.1113/jphysiol.2009.179432)
- 614 48. Sallam N, Laher I. 2015 Exercise modulates oxidative stress and inflammation in aging and
615 cardiovascular diseases. *Oxidative Medicine and Cellular Longevity* **2016**, e7239639.
616 (doi:10.1155/2016/7239639)
- 617 49. Hochachka PW, Somero GN. 2002 *Biochemical Adaptation: Mechanism and Process in*
618 *Physiological Evolution*. Oxford University Press.
- 619 50. Teulier L, Omlin T, Weber J-M. 2013 Lactate kinetics of rainbow trout during graded exercise:
620 do catheters affect the cost of transport? *Journal of Experimental Biology* **216**, 4549–4556.
621 (doi:10.1242/jeb.091058)
- 622 51. Zhang Y, Lauder GV. 2023 Energetics of collective movement in vertebrates. *Journal of*
623 *Experimental Biology* **226**, jeb245617. (doi:10.1242/jeb.245617)
- 624 52. Zhang Y, Lauder GV. 2024 Energy conservation by collective movement in schooling fish. *eLife*
625 **12**. (doi:10.7554/eLife.90352.2)
- 626 53. Zhang Y, Lauder GV. 2025 Physics and physiology of fish collective movement. *Newton* **1**.
627 (doi:10.1016/j.newton.2025.100021)
- 628 54. Hochachka PW, Mommsen TP. 1983 Protons and anaerobiosis. *Science* **219**, 1391–1397.
629 (doi:10.1126/science.6298937)
- 630 55. Hochachka PW, Buck LT, Doll CJ, Land SC. 1996 Unifying theory of hypoxia tolerance:
631 molecular/metabolic defense and rescue mechanisms for surviving oxygen lack. *Proc Natl Acad*
632 *Sci U S A* **93**, 9493–9498.
- 633 56. Gatten RE, Caldwell JP, Stockard ME. 1984 Anaerobic metabolism during intense swimming by
634 anuran larvae. *Herpetologica* **40**, 164–169.
- 635 57. Wieser W, Platzner U, Hinterleitner S. 1985 Anaerobic and aerobic energy production of young
636 rainbow trout (*Salmo gairdneri*) during and after bursts of activity. *J Comp Physiol B* **155**, 485–
637 492. (doi:10.1007/BF00684679)

- 638 58. Marras S, Claireaux G, McKenzie DJ, Nelson JA. 2010 Individual variation and repeatability in
639 aerobic and anaerobic swimming performance of European sea bass, *Dicentrarchus labrax*.
640 *Journal of Experimental Biology* **213**, 26–32. (doi:10.1242/jeb.032136)
- 641 59. Burnett NJ, Hinch SG, Braun DC, Casselman MT, Middleton CT, Wilson SM, Cooke SJ. 2014
642 Burst Swimming in areas of high flow: Delayed consequences of anaerobiosis in wild adult
643 sockeye salmon. *Physiological and Biochemical Zoology* **87**, 587–598. (doi:10.1086/677219)
- 644 60. Garenc C, Couture P, Laflamme M-A, Guderley H. 1999 Metabolic correlates of burst swimming
645 capacity of juvenile and adult threespine stickleback (*Gasterosteus aculeatus*). *J Comp Physiol B*
646 **169**, 113–122. (doi:10.1007/s003600050201)
- 647 61. Beamish FWH. 1978 Swimming capacity. In *Fish Physiology*, pp. 101–187. London: Academic
648 Press.
- 649 62. Jones DR, Randall DJ. 1978 The respiratory and circulatory systems during exercise in fish. *Fish*
650 *Physiology* **7**, 425–501.
- 651 63. Steffensen JF. 1989 Some errors in respirometry of aquatic breathers: How to avoid and
652 correct for them. *Fish Physiol Biochem* **6**, 49–59. (doi:10.1007/BF02995809)
- 653 64. Chabot D, Claireaux G. 2008 Environmental hypoxia as a metabolic constraint on fish: The case
654 of Atlantic cod, *Gadus morhua*. *Marine Pollution Bulletin* **57**, 287–294.
655 (doi:10.1016/j.marpolbul.2008.04.001)
- 656 65. Svendsen MBS, Bushnell PG, Steffensen JF. 2016 Design and setup of intermittent-flow
657 respirometry system for aquatic organisms. *Journal of Fish Biology* **88**, 26–50.
658 (doi:10.1111/jfb.12797)
- 659 66. Chabot D, Steffensen JF, Farrell AP. 2016 The determination of standard metabolic rate in
660 fishes. *Journal of Fish Biology* **88**, 81–121. (doi:10.1111/jfb.12845)
- 661 67. Wood CM. 2018 The fallacy of the Pcrit – are there more useful alternatives? *Journal of*
662 *Experimental Biology* **221**, jeb163717. (doi:10.1242/jeb.163717)
- 663 68. Chabot D, Zhang Y, Farrell AP. 2021 Valid oxygen uptake measurements: using high r² values
664 with good intentions can bias upward the determination of standard metabolic rate. *Journal of*
665 *Fish Biology* **98**, 1206–1216. (doi:10.1111/jfb.14650)
- 666 69. Zhang Y, Gilbert MJH, Farrell AP. 2020 Measuring maximum oxygen uptake with an
667 incremental swimming test and by chasing rainbow trout to exhaustion inside a respirometry
668 chamber yields the same results. *Journal of Fish Biology* **97**, 28–38. (doi:10.1111/jfb.14311)
- 669 70. Prinzing TS, Zhang Y, Wegner NC, Dulvy NK. 2021 Analytical methods matter too: Establishing a
670 framework for estimating maximum metabolic rate for fishes. *Ecology and Evolution* **11**(15),
671 9987–10003. (doi:10.1002/ece3.7732)
- 672 71. Zhang Y, Montgomery DW, White CF, Richards JG, Brauner CJ, Farrell AP. 2022 Characterizing
673 the hypoxic performance of a fish using a new metric: PAAS-50. *Journal of Experimental*
674 *Biology* , jeb.244239. (doi:10.1242/jeb.244239)
- 675 72. Zhang Y, Farrell AP. 2022 Testing the hypoxia tolerance and hypoxic performance of fishes: A
676 two-tier screening approach. *Frontiers in Marine Science* **9**, 939239.

- 677 73. Near TJ *et al.* 2013 Phylogeny and tempo of diversification in the superradiation of spiny-rayed
678 fishes. *Proceedings of the National Academy of Sciences* **110**, 12738–12743.
679 (doi:10.1073/pnas.1304661110)
- 680 74. Wainwright PC, Longo SJ. 2017 Functional Innovations and the Conquest of the Oceans by
681 Acanthomorph Fishes. *Current Biology* **27**, R550–R557. (doi:10.1016/j.cub.2017.03.044)
- 682 75. Andersen Ø, Wetten OF, De Rosa MC, Andre C, Carelli Alinovi C, Colafranceschi M, Brix O,
683 Colosimo A. 2008 Haemoglobin polymorphisms affect the oxygen-binding properties in Atlantic
684 cod populations. *Proceedings of the Royal Society B: Biological Sciences* **276**, 833–841.
685 (doi:10.1098/rspb.2008.1529)
- 686 76. Wells RMG. 2009 Chapter 6 Blood-gas transport and hemoglobin function: Adaptations for
687 functional and environmental hypoxia. In *Fish Physiology* (eds JG Richards, AP Farrell, CJ
688 Brauner), pp. 255–299. Academic Press. (doi:10.1016/S1546-5098(08)00006-X)
- 689 77. Sidell BD, O'Brien KM. 2006 When bad things happen to good fish: the loss of hemoglobin and
690 myoglobin expression in Antarctic icefishes. *Journal of Experimental Biology* **209**, 1791–1802.
691 (doi:10.1242/jeb.02091)
- 692 78. Pan YK, Ern R, Morrison PR, Brauner CJ, Esbaugh AJ. 2017 Acclimation to prolonged hypoxia
693 alters hemoglobin isoform expression and increases hemoglobin oxygen affinity and aerobic
694 performance in a marine fish. *Sci Rep* **7**, 7834. (doi:10.1038/s41598-017-07696-6)
- 695 79. Brauner CJ, Val AL, Randall DJ. 1993 The effect of graded methaemoglobin levels on the
696 swimming performance of chinook salmon (*Oncorhynchus tshawytscha*). *Journal of*
697 *Experimental Biology* **185**, 121–135. (doi:10.1242/jeb.185.1.121)
- 698 80. Nikinmaa M, Berenbrink M, Brauner CJ. 2019 Regulation of erythrocyte function: Multiple
699 evolutionary solutions for respiratory gas transport and its regulation in fish. *Acta Physiologica*
700 **227**, e13299. (doi:10.1111/apha.13299)
- 701 81. Göpel T, Burggren WW. 2022 Insufficient reporting of experimental variables as a cause for
702 nonreproducibility in animal physiology? A case study. *American Journal of Physiology-*
703 *Regulatory, Integrative and Comparative Physiology* **323**, R363–R374.
704 (doi:10.1152/ajpregu.00026.2022)
- 705 82. Williams TD. 2007 Individual variation in endocrine systems: moving beyond the 'tyranny of the
706 Golden Mean'. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 1687–
707 1698. (doi:10.1098/rstb.2007.0003)
- 708 83. Wergel-Kolmert U, Wisén A, Wohlfart B. 2002 Repeatability of measurements of oxygen
709 consumption, heart rate and Borg's scale in men during ergometer cycling. *Clinical Physiology*
710 *and Functional Imaging* **22**, 261–265. (doi:10.1046/j.1475-097x.2002.00428.x)
- 711 84. Kolok AS. 1992 The swimming performances of individual largemouth bass (*Micropterus*
712 *Salmoides*) are repeatable. *Journal of Experimental Biology* **170**, 265–270.
713 (doi:10.1242/jeb.170.1.265)
- 714 85. Nelson JA, Claireaux G. 2005 Sprint swimming performance of juvenile european sea bass.
715 *Transactions of the American Fisheries Society* **134**, 1274–1284. (doi:10.1577/T04-087.1)

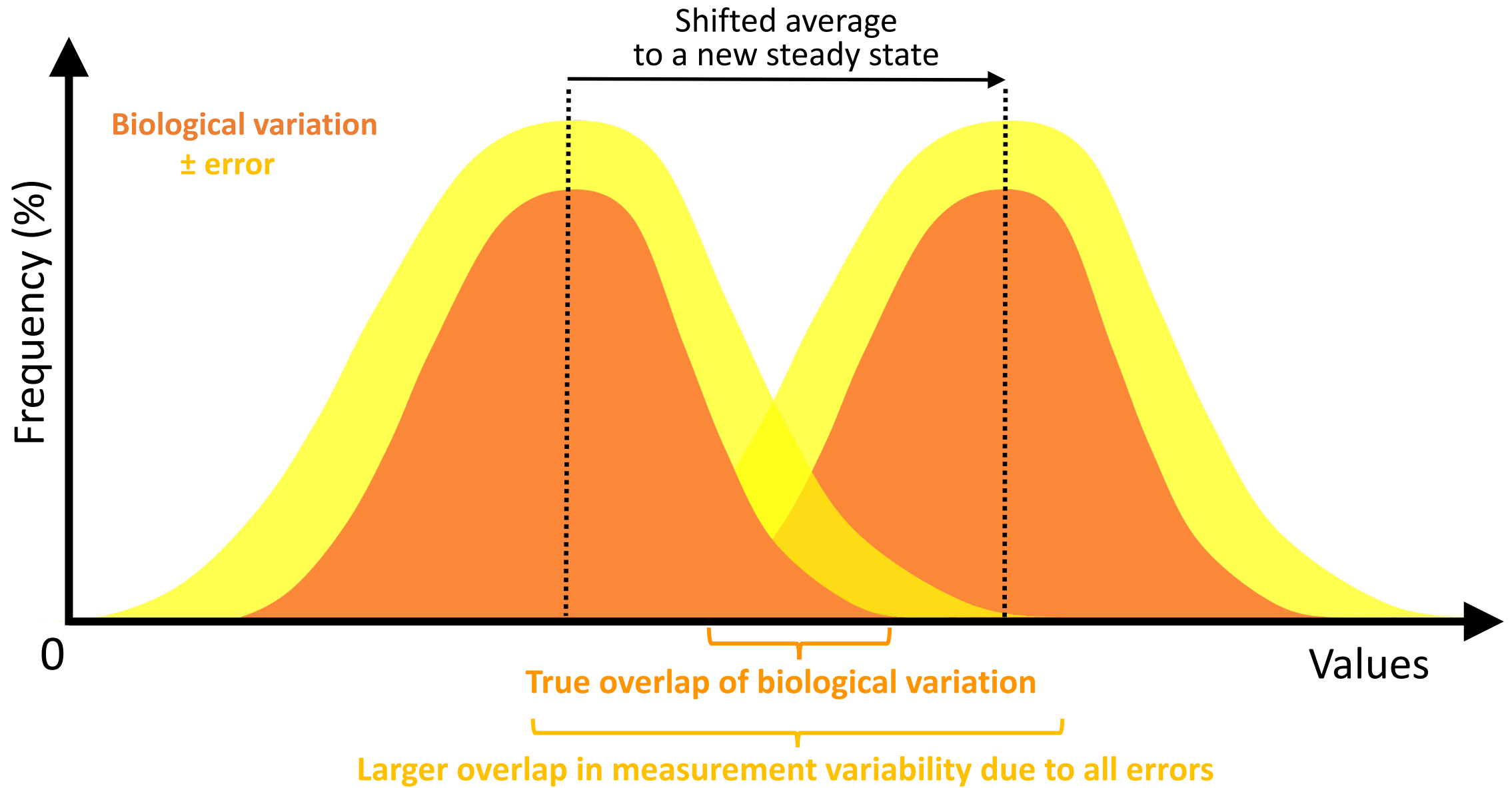
- 716 86. Claireaux G, Handelsman C, Standen E, Nelson JA. 2007 Thermal and temporal stability of
717 swimming performance in the European sea bass. *Physiological and Biochemical Zoology* **80**,
718 186–196. (doi:10.1086/511143)
- 719 87. Nelson JA, Gotwalt PS, Simonetti CA, Snodgrass JW. 2008 Environmental correlates, plasticity,
720 and repeatability of differences in performance among blacknose dace (*Rhinichthys atratulus*)
721 populations across a gradient of urbanization. *Physiological and Biochemical Zoology* **81**, 25–
722 42. (doi:10.1086/523304)
- 723 88. Nelson JA, Atzori F, Gastrich KR. 2015 Repeatability and phenotypic plasticity of fish swimming
724 performance across a gradient of urbanization. *Environ Biol Fish* **98**, 1431–1447.
725 (doi:10.1007/s10641-014-0369-x)
- 726 89. Norin T, Rowsey LE, Houslay TM, Reeve C, Speers-Roesch B. 2024 Among-individual variation in
727 thermal plasticity of fish metabolic rates causes profound variation in temperature-specific
728 trait repeatability, but does not co-vary with behavioural plasticity. *Philosophical Transactions*
729 *of the Royal Society B: Biological Sciences* **379**, 20220488. (doi:10.1098/rstb.2022.0488)
- 730 90. Nakagawa S, Schielzeth H. 2010 Repeatability for gaussian and non-gaussian data: a practical
731 guide for biologists. *Biol Rev Camb Philos Soc* **85**, 935–956. (doi:10.1111/j.1469-
732 185X.2010.00141.x)
- 733 91. Dingemanse NJ, Dochtermann NA. 2013 Quantifying individual variation in behaviour: mixed-
734 effect modelling approaches. *Journal of Animal Ecology* **82**, 39–54. (doi:10.1111/1365-
735 2656.12013)
- 736 92. Schielzeth H, Nakagawa S. 2022 Conditional repeatability and the variance explained by
737 reaction norm variation in random slope models. *Methods in Ecology and Evolution* **13**, 1214–
738 1223. (doi:10.1111/2041-210X.13856)
- 739 93. Kar F, Nakagawa S, Friesen CR, Noble DWA. 2021 Individual variation in thermal plasticity and
740 its impact on mass-scaling. *Oikos* **130**, 1131–1142. (doi:10.1111/oik.08122)
- 741 94. Kar F, Nakagawa S, Noble DWA. 2022 Impact of developmental temperatures on thermal
742 plasticity and repeatability of metabolic rate. *Evol Ecol* **36**, 199–216. (doi:10.1007/s10682-022-
743 10160-1)
- 744 95. Altman DG, Bland JM. 1983 Measurement in medicine: The analysis of method comparison
745 studies. *Journal of the Royal Statistical Society Series D: The Statistician* **32**, 307–317.
746 (doi:10.2307/2987937)
- 747 96. Bland JM, Altman DG. 2003 Applying the right statistics: analyses of measurement studies.
748 *Ultrasound Obstet Gynecol* **22**, 85–93. (doi:10.1002/uog.122)
- 749 97. Sokal RR, Rohlf FJ. 1995 *Biometry*. Macmillan.
- 750 98. Zhang Y, Polinski MP, Morrison PR, Brauner CJ, Farrell AP, Garver KA. 2019 High-load reovirus
751 infections do not imply physiological impairment in salmon. *Front. Physiol.* **10**.
752 (doi:10.3389/fphys.2019.00114)
- 753 99. Polinski MP, Zhang Y, Morrison PR, Marty GD, Brauner CJ, Farrell AP, Garver KA. 2021 Innate
754 antiviral defense demonstrates high energetic efficiency in a bony fish. *BMC Biology* **19**, 138.
755 (doi:10.1186/s12915-021-01069-2)

100. Zhang Y. 2021 Interpreting species, intraspecific and intra-individual variability by comprehensively characterizing a fish's respiratory phenotype with valid measures of oxygen uptake. University of British Columbia. (doi:10.14288/1.0396683)
101. Fangue NA, Podrabsky JE, Crawshaw LI, Schulte PM. 2009 Countergradient variation in temperature preference in populations of killifish *Fundulus heteroclitus*. *Physiological and Biochemical Zoology: Ecological and Evolutionary Approaches* **82**, 776–786. (doi:10.1086/606030)
102. Sandblom E, Gräns A, Axelsson M, Seth H. 2014 Temperature acclimation rate of aerobic scope and feeding metabolism in fishes: implications in a thermally extreme future. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20141490. (doi:10.1098/rspb.2014.1490)
103. Moyano M, Candebat C, Ruhbaum Y, Álvarez-Fernández S, Claireaux G, Zambonino-Infante J-L, Peck MA. 2017 Effects of warming rate, acclimation temperature and ontogeny on the critical thermal maximum of temperate marine fish larvae. *PLOS ONE* **12**, e0179928. (doi:10.1371/journal.pone.0179928)
104. Rees BB *et al.* 2024 Estimating maximum oxygen uptake of fishes during swimming and following exhaustive chase – different results, biological bases and applications. *Journal of Experimental Biology* **227**, jeb246439. (doi:10.1242/jeb.246439)
105. Scott GR, Milsom WK. 2006 Flying high: a theoretical analysis of the factors limiting exercise performance in birds at altitude. *Respir Physiol Neurobiol* **154**, 284–301. (doi:10.1016/j.resp.2006.02.012)
106. Bassett DR, Howley ET. 2000 Limiting factors for maximum oxygen uptake and determinants of endurance performance. *Medicine & Science in Sports & Exercise* **32**, 70–84. (doi:10.1097/00005768-200001000-00012)
107. Cooper S-M, Baker JS, Tong RJ, Roberts E, Hanford M. 2005 The repeatability and criterion related validity of the 20 m multistage fitness test as a predictor of maximal oxygen uptake in active young men. *British Journal of Sports Medicine* **39**, e19. (doi:10.1136/bjsm.2004.013078)
108. Claireaux G, Lefrançois C. 2007 Linking environmental variability and fish performance: integration through the concept of scope for activity. *Philosophical Transactions of the Royal Society B* **362**, 2031–2041. (doi:10.1098/rstb.2007.2099)
109. Eliason EJ *et al.* 2011 Differences in thermal tolerance among sockeye salmon populations. *Science* **332**, 109–112. (doi:10.1126/science.1199158)
110. Anttila K, Couturier CS, Øverli Ø, Johnsen A, Marthinsen G, Nilsson GE, Farrell AP. 2014 Atlantic salmon show capability for cardiac acclimation to warm temperatures. *Nature Communications* **5**, 4252. (doi:10.1038/ncomms5252)
111. Auer SK, Dick CA, Metcalfe NB, Reznick DN. 2018 Metabolic rate evolves rapidly and in parallel with the pace of life history. *Nature Communications* **9**, 14. (doi:10.1038/s41467-017-02514-z)
112. Jutfelt F *et al.* 2018 Oxygen- and capacity-limited thermal tolerance: blurring ecology and physiology. *Journal of Experimental Biology* **221**, jeb169615. (doi:10.1242/jeb.169615)
113. Gaesser GA, Brooks GA. 1984 Metabolic bases of excess post-exercise oxygen consumption: a review. *Medicine & Science in Sports & Exercise* **16**, 29–43.

- 796 114. Wood CM. 1991 Acid-base and ion balance, metabolism, and their Interactions, after
797 exhaustive exercise in Fish. *Journal of Experimental Biology* **160**, 285–308.
798 (doi:10.1242/jeb.160.1.285)
- 799 115. Zhang Y, Claireaux G, Takle H, Jørgensen SM, Farrell AP. 2018 A three-phase excess post-
800 exercise oxygen consumption in Atlantic salmon *Salmo salar* and its response to exercise
801 training. *Journal of Fish Biology* **92**, 1385–1403. (doi:10.1111/jfb.13593)
- 802 116. Laforgia J, Withers RT, Gore CJ. 2006 Effects of exercise intensity and duration on the excess
803 post-exercise oxygen consumption. *Journal of Sports Sciences* **24**, 1247–1264.
804 (doi:10.1080/02640410600552064)
- 805 117. Zhang Y, So BE, Farrell AP. 2021 Hypoxia performance curve: Assess a whole-organism
806 metabolic shift from a maximum aerobic capacity towards a glycolytic capacity in fish.
807 *Metabolites* **11**, 447. (doi:10.3390/metabo11070447)
- 808 118. Smith RL, Soeters MR, Wüst RCI, Houtkooper RH. 2018 Metabolic flexibility as an adaptation to
809 energy resources and requirements in health and disease. *Endocrine Reviews* **39**, 489–517.
810 (doi:10.1210/er.2017-00211)
- 811 119. Brooks GA. 2018 The science and translation of lactate shuttle theory. *Cell Metabolism* **27**,
812 757–785. (doi:10.1016/j.cmet.2018.03.008)
- 813 120. Kelley DE, Reilly JP, Veneman T, Mandarino LJ. 1990 Effects of insulin on skeletal muscle
814 glucose storage, oxidation, and glycolysis in humans. *American Journal of Physiology-
815 Endocrinology and Metabolism* (doi:10.1152/ajpendo.1990.258.6.E923)
- 816 121. Nelson JA, Kraskura K, Lipkey GK. 2019 Repeatability of hypoxia tolerance of individual juvenile
817 striped bass *morone saxatilis* and effects of social status. *Physiological and Biochemical
818 Zoology* **92**, 396–407. (doi:10.1086/704010)
- 819 122. Clark TD, Raby GD, Roche DG, Binning SA, Speers-Roesch B, Jutfelt F, Sundin J. 2020 Ocean
820 acidification does not impair the behaviour of coral reef fishes. *Nature* **577**, 370–375.
821 (doi:10.1038/s41586-019-1903-y)
- 822 123. Clark TD, Raby GD, Roche DG, Binning SA, Speers-Roesch B, Jutfelt F, Sundin J. 2020 Reply to:
823 Methods matter in repeating ocean acidification studies. *Nature* **586**, E25–E27.
824 (doi:10.1038/s41586-020-2804-9)
- 825 124. Noble DWA, Pottier P, Lagisz M, Burke S, Drobniak SM, O’Dea RE, Nakagawa S. 2022 Meta-
826 analytic approaches and effect sizes to account for ‘nuisance heterogeneity’ in comparative
827 physiology. *Journal of Experimental Biology* **225**, jeb243225. (doi:10.1242/jeb.243225)
- 828 125. Bothe TL, Pilz N, Patzak A, Opatz OS. 2023 Bridging the gap: The dichotomy between
829 measurement and reality in physiological research. *Acta Physiologica* **238**, e14015.
830 (doi:10.1111/apha.14015)
- 831 126. Atkinson G, Nevill AM. 1998 Statistical methods for assessing measurement error (reliability) in
832 variables relevant to sports medicine. *Sports Med* **26**, 217–238. (doi:10.2165/00007256-
833 199826040-00002)

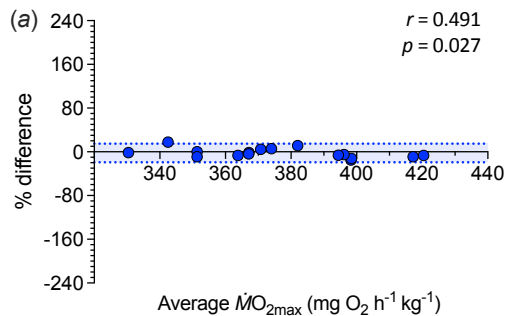
- 834 127. Martínez M, Guderley H, Nelson JA, Webber D, Dutil J -D. 2002 Once a fast cod, always a fast
835 cod: Maintenance of performance hierarchies despite changing food availability in cod (*Gadus*
836 *morhua*). *Physiological and Biochemical Zoology* **75**, 90–100. (doi:10.1086/339213)
- 837 128. Wolak ME, Fairbairn DJ, Paulsen YR. 2012 Guidelines for estimating repeatability. *Methods in*
838 *Ecology and Evolution* **3**, 129–137. (doi:10.1111/j.2041-210X.2011.00125.x)

839

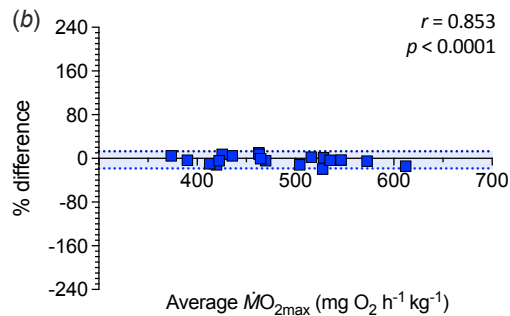


Repeatable metric

S. salar :



D. labrax :



Unrepeatable metric

