

Metrics for optimum allocation of resources in the composition and characterization of crop collections: The CIMMYT wheat collection as a proof of concept

M. Humberto Reyes-Valdés¹, Juan Burgueño^{2*}, Carolina Paola Sansaloni², Thomas Payne², Angela Pacheco² and Areli González-Cortés³.

Supporting Information

Supporting Information Mathematical addendum

The Kullback-Leibler divergence as a criterion for relative balance

As stated in the paper, let a_i and o_i be the actual and ideal proportions for the end-groups in a hierarchical tree for a crop collection, with $o_i > 0$ for all i . Define the Kullback-Leibler divergence of the ensemble of actual proportions from the optimal ones as follows:

$$D = \sum_{i=1}^n a_i \log_2 \frac{a_i}{o_i}$$

The formula for D can be expressed as follows:

$$\begin{aligned} D &= \sum_{i=1}^n a_i [\log_2(a_i) - \log_2(o_i)] \\ &= \sum_{i=1}^n a_i \log_2(o_i) + \sum_{i=1}^n a_i \log_2(a_i) \\ &= - \sum_{i=1}^n a_i \log_2(o_i) + \sum_{i=1}^n a_i \log_2(a_i) \end{aligned}$$

The second term of the last expression is the negative Shannon entropy of the vector a of n actual proportions in the collection. Thus:

$$D = - \sum_{i=1}^n a_i \log_2(o_i) - H(a)$$

The first term in the last expression is the expectation of $-\log_2(o)$, weighted by the respective a_i values. Thus, for a fixed vector of optimal proportions o , the maximum value of such an expectation occurs when a value o_x , such that $o_x \leq o_i$ for all i , is weighted by 1 and the remaining o values are weighted by 0. In such a case, the maximum of the first term is attained and equals $-\log_2(\min(o))$. Now, for the second term, a set of a_i values with one of them being 1 and the remaining being 0, gives $H(a) = 0$, which is the minimum for the Shannon entropy. Since the established condition for the vector a maximizes the first term and minimizes $H(a)$, then it maximizes D . Let us call u a vector of proportions that fulfills the given conditions; then, the maximum value of D is:

$$\begin{aligned} \max(D) &= -\log_2(\min(o)) - H(u) \\ &= -\log_2(\min(o)), \end{aligned}$$

since $H(u) = 0$.

The divergence D can be divided by its maximum to have a relative value $D / -\log_2(\min(o))$. The Relative Balance, defined in Equation 3, is one minus the relative D value. Thus:

$$\begin{aligned} RB &= 1 - \frac{D}{-\log_2(\min(o))} \\ &= 1 + \frac{D}{\log_2(\min(o))} \end{aligned}$$

Supporting Information Definition of geographic areas

The geographic zones were grouped as follows:

AMERICA Antigua and Barbuda, Argentina, Bahamas, Barbados, Belize, Bolivarian Republic of Venezuela, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, El Salvador, French Guiana, Grenada, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and Grenadines, Suriname, Trinidad and Tobago, United States of America, Uruguay.

ASIA This group excludes China and CWANA countries. It includes: Bahrain, Bangladesh, Bhutan, Brunei Darussalam, Cambodia, Cyprus, India, Indonesia, Israel, Japan, Korea North, Korea South, Kuwait, Lao PDR, Malaysia, Maldives, Mongolia, Myanmar, Nepal, Philippines, Qatar, Republic of China Taiwan, Singapore, Sri Lanka, Thailand, Viet Nam.

BEUROPE Balkans and eastern Europe. It excludes countries of western Europe. It includes: Albania, Belarus, Bulgaria, Czech Republic, Hungary, Moldova, Montenegro, Poland, Republic of Macedonia, Romania, Russian Federation, Slovakia, Ukraine.

CHINA All accessions with ORIGCTY coded as CHINA.

CWANA Central and West Asia and North Africa. Afghanistan, Algeria, Armenia, Azerbaijan, Belgium, Egypt, Ethiopia, Georgia, Iraq, Islamic Republic of Iran, Jordan, Kazakhstan, Kyrgyzstan, Lebanon, Libya, Mauritania, Morocco, Oman, Pakistan, Palestinian Territory, Saudi Arabia, Somalia, Sudan, Syrian Arab Republic, Tajikistan, Tunisia, Turkey, Turkmenistan, United Arab Emirates, Uzbekistan, Yemen.

OCEANIA All Oceania countries, with the exception of those belonging to America. It includes: Australia, Federal States of Micronesia, Fiji, Indonesia, Kiribati, Marshall Islands, Nauru, New Zealand, Palau, Papua New Guinea, Samoa, Solomon Islands, Tonga, Tuvalu, Vanuatu.

WEUROPE Western Europe, with the exception of those countries classified in CWANA: Andorra, Austria, Bosnia and Herzegovina, Croatia, Denmark, Faroe Islands, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Liechtenstein, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, San Marino, Spain, Sweden, Switzerland, United Kingdom.

OTHER All countries not classified in the above described groups.