# **scientific** reports

Check for updates

OPEN

# PeruNPDB: the Peruvian Natural Products Database for in silico drug screening

Haruna L. Barazorda-Ccahuana[1], Lena Gálvez Ranilla[2,3], Mayron Antonio Candia-Puma[1,4], Eymi Gladys Cárcamo-Rodriguez[1,4], Angela Emperatriz Centeno-Lopez[1,4], Gonzalo Davila-Del-Carpio[4], José L. Medina-Franco[5] & Miguel Angel Chávez-Fumagalli[1✉]

Since the number of drugs based on natural products (NPs) represents a large source of novel pharmacological entities, NPs have acquired significance in drug discovery. Peru is considered a megadiverse country with many endemic species of plants, terrestrial, and marine animals, and microorganisms. NPs databases have a major impact on drug discovery development. For this reason, several countries such as Mexico, Brazil, India, and China have initiatives to assemble and maintain NPs databases that are representative of their diversity and ethnopharmacological usage. We describe the assembly, curation, and chemoinformatic evaluation of the content and coverage in chemical space, as well as the physicochemical attributes and chemical diversity of the initial version of the Peruvian Natural Products Database (PeruNPDB), which contains 280 natural products. Access to PeruNPDB is available for free (https://perunpdb.com.pe/). The PeruNPDB's collection is intended to be used in a variety of tasks, such as virtual screening campaigns against various disease targets or biological endpoints. This emphasizes the significance of biodiversity protection both directly and indirectly on human health.

Biodiversity is the variety of all life forms, including the morphological diversity of individuals and populations within a species, the taxonomic diversity of species within a community or ecosystem, the functional diversity of groups of species within an ecosystem, and the diversity of ecosystems themselves[1]. While the total number of species in every taxonomic group has been predicted for all kingdoms of life on earth at approximately 8.7 million[2,3], it is remarkable that the distribution of that vast number of species is highly concentrated in specific areas. These regions are particularly important for biodiversity conservation and are called biodiversity hotspots[4], although: Bolivia, Brazil, China, Colombia, Costa Rica, the Democratic Republic of Congo, Ecuador, India, Indonesia, Kenya, Madagascar, Malaysia, Mexico, Peru, Philippines, South Africa, and Venezuela are considered megadiverse countries[5]. Peru occupies the seventh place in this group, as it possesses 28 of the 32 existing climates in the world and 84 of the 103 life zones known on earth. This is evidenced by considering that the country has 25,000 plant species or 10% of the entire number of species worldwide, whereas 30% are endemic, and endemic animal species such as 115 birds, 109 mammals, and 185 amphibians species, which represent 6, 27.5 and 48.5% of the total number worldwide, respectively[6,7].

Biodiversity conservation is important since plants, animals, and other life forms such as bacteria, archaea, protozoa, and fungi, are used directly or indirectly to produce pharmaceuticals, and for their scientific value, among other resources[8]. The number of drugs derived from natural products (NP) that were introduced to the market over forty years represented a significant source of new pharmacological entities[9]. Whilst the Peruvian population uses approximately 5000 Peruvian plants for 49 purposes or applications, where about 1400 species are described as medicinal[10–13]. The contribution from traditional Peruvian medicine can be embodied by Quinine, a component of the bark of the cinchona tree (*Cinchona officinalis*), employed in the treatment of Malaria[14]. Additionally, two other valuable contributions to modern pharmacopeias such as the coca plant (*Erythroxylum coca*), from which cocaine was first isolated and later led to local anesthetics[15], and the balsam of Peru (*Myroxylon*

[1]Computational Biology and Chemistry Research Group, Vicerrectorado de Investigación, Universidad Católica de Santa María, 04000 Arequipa, Peru. [2]Laboratory of Research in Food Science, Universidad Catolica de Santa Maria, 04000 Arequipa, Peru. [3]Escuela Profesional de Ingeniería de Industria Alimentaria, Facultad de Ciencias e Ingenierías Biológicas y Químicas, Universidad Catolica de Santa Maria, 04000 Arequipa, Peru. [4]Facultad de Ciencias Farmacéuticas, Bioquímicas y Biotecnológicas, Universidad Católica de Santa María, 04000 Arequipa, Peru. [5]DIFACQUIM Research Group, Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico. ✉email: mchavezf@ucsm.edu.pe

*balsamum*), which was used wide-reaching for the treatment of wounds[16], can be mentioned. However, the potential of Peruvian NPs remains underexploited since most of these useful native species can be domesticated or semi-domesticated[17]. Also, the amount and nature of experimental evidence published on active NPs are still limited[18], and most of the current studies reported crude medicinal activities, while potentially active NPs have been isolated only from a few numbers of plants[19].

Computer-aided drug design (CADD), one of the key approaches to modern pre-clinical drug discovery, can be defined as computational methods that are applied to discover, develop, and analyze drugs and active molecules[20]. Among the key approaches that comprise CADD, virtual screening is one of the major contributors to CADD since it stands as a contemporary approach to the experimental in vitro high-throughput screening (HTS) for hit identification and optimization[21]. Integrating CADD approaches to curated databases, which are described as a well-organized collection of data in any field, the drug development process may be sped up and cost reduced[22]. Considering this, large databases containing NPs from various data sources have been released, such as the COlleCtion of Open Natural prodUcTs (COCONUT), which contains 406,076 unique "flat" NPs, and a total of 730,441 NPs where stereochemistry has been preserved[23]; and the LOTUS initiative, which has 750,000 referenced structure-organism pairs[24]. Also, several NPs compound databases from particular geographical locations have been assembled, such as the Traditional Chinese Medicine (TCM) Database@Taiwan database containing approximately 58,000 molecules[25]; the Indian Medicinal Plants, Phytochemistry and Therapeutics 2.0 (IMPPAT 2.0) which contains more than 10,000 phytochemicals[26]; and the AfroDB which is composed of around 1000 NPs[27]. Likewise, some countries in Latin America have published their own public NPs databases such as NuBBEDB which contains more than 2000 NPs[28], and SistematX which contains more than 2500 NPs[29], both from Brazil, and BIOFACQUIM from Mexico, which contains a total of 531 molecules[30]. Furthermore, NPs databases had been used as a repository to identify several promising candidates to be considered for further development for the treatment of diseases[31], such as Chagas disease[32,33], Tuberculosis[34], Leishmaniasis[35,36], Schistosomiasis[37], and COVID-19[38]. The present work introduces the first version of the Peruvian Natural Products Database (PeruNPDB), describing its assembly, curation, and chemoinformatic characterization of molecular diversity and coverage in chemical space. The database is freely available at the web-interface PeruNPDB Explorer (https://perunpdb.com.pe/). We anticipate that the PeruNPDB will make it possible to conduct additional virtual screening tests to create innovative pharmacological entities and other biotechnological approaches and serve as a resource for information on conservation guidelines.

## Methods

### Search strategy, study selection, and data extraction.
A systematic review search strategy to examine the literature for studies describing NP from Peruvian sources was adapted from[39]. Whereas PubMed, the main database for the health sciences, maintained by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM), is a database that contains about 32 million citations, belonging to more than 5300 journals currently indexed in MEDLINE[40]; it provides uniform indexing of biomedical literature, the Medical Subject Headings (MeSH terms), which form a controlled vocabulary or specific set of terms that describe the topic of a paper consistently and uniformly[41]. Firstly, to find terms associated in the literature with Peruvian NPs, the MeSH terms "*Peru*" AND the "*Natural Products*" were employed in a search carried out at the PubMed database (https://pubmed.ncbi.nlm.nih.gov/), (last searched on 10 June 2022), though the results were plotted into a network map of the co-occurrence of MeSH terms in the VOSviewer software (version 1.6.17)[42], which employs a modularity-based method algorithm to measure the strength of clusters[43]. The resultant cluster content was analyzed to select relevant studies associated with Peruvian NPs. Three phases went into selecting the studies. First, papers written in languages other than English, copies of articles, reviews, and meta-analyses were disregarded. The highly relevant full studies were then retrieved and separated from the papers with a title or abstract that did not provide enough information to be included. Next, the titles and abstracts of the publications chosen through the search approach were visually evaluated. The data supplied from each investigation contained the NP's characterization as well as details on the genus and species of the sources from which the NP were isolated. Additionally, the information from the bibliographic reference was extracted, even if all research that discussed chemicals derived from Peruvian natural sources was already considered.

### PeruNPDB assemble and molecular properties calculation.
The simplified molecular-input line-entry system (SMILES)[44] of compounds previously described in the NPs selected in the previous step were searched and retrieved from PubChem[45], DrugBank[46], or ChEMBL[47] servers, while for unavailable NPs the ChemDraw tool[48] was employed to generate the SMILE notation. Moreover, the Osiris DataWarrior v05.02.01 software[49] was employed to generate the dataset's structure data files (SDFs). This followed the uploading to the Konstanz information miner (KNIME) Analytics Platform[50], where the "Molecular Type Cast", and the "RDKit Structure Normalizer" KNIME nodes were employed to curate the chemical structures on the dataset. Moreover, for every compound in the dataset, the classification system for describing small molecule structures is described based on NP Classifier[51], which employs a biosynthetic ontology that is specific to natural products; or ClassyFire[52] which is a general classification system for small molecules that are based on the ChemOnt ontology, was employed. The KNIME's "RKDit Descriptor Calculator" node was employed to calculate six physico-chemical properties of therapeutic interest, namely: molecular weight (MW), octanol/water partition coefficient (clogP), topological surface area (TPSA), aqueous solubility (clogS), number of H-bond donor atoms (HBD) and number of H-bond acceptor atoms (HBA) of the PeruNPDB, while the statistical analysis was done within the GraphPad Prism software version 9.4.0 for Windows, GraphPad Software, San Diego, California USA, http://www.graphpad.com, by calculating the mean, median, standard deviation, and the coefficient of variation of the calculated properties. Box-and-whisker plots showing, the maximum and minimum values were generated

for visualization, and the One-way ANOVA followed by Dunnett correction for multiple comparisons test was employed to evaluate the differences between the datasets. The results were considered statistically significant when $p<0.05$.

**Visual representation of chemical space.** To generate a visual representation of the chemical space of the PeruNPDB, two visualization methods, for the auto-scaled six properties of pharmaceutical interest, namely: MW, ClogP, TPSA, clogS, HBD, and HBA, were employed: principal component analysis (PCA), which reduces data dimensions by geometrically projecting them onto lower dimensions called principal components (PCs)[53] calculated by the "PCA" KNIME node. The second technique was the t-distributed stochastic neighbor embedding (t-SNE), which is a nonlinear dimension reduction in which Gaussian probability distributions over high-dimensional space are constructed and used to optimize a student t-distribution in low-dimensional space[54], calculated by the t-SNE (L. Jonsson) KNIME node. Three and two-dimensional scatter-plot representations were generated for PCA and t-SNE, respectively with the Plotly KNIME node. Additionally, the Tanimoto similarity score was calculated for clustering the compounds, while the atom-pair-based fingerprints of the NPs were obtained using the "ChemmineR" package[55] in the R programming environment (version 4.0.3)[56], a heatmap was generated for visualization. The same procedure was employed in the reference datasets: AfroDB[27], BIOFAQUIM[30], and NUBBEDB[28] retrieved from the ZINC20 database[57].

**Global diversity: consensus diversity analysis.** Since chemical diversity strongly depends on the structure representation, it is reasonable to consider multiple representations for a complete global assessment. The consensus diversity (CD) plots have been proposed as simple two-dimensional graphs that enable the comparison of the diversity of compound data sets using four sets of structural representations: the molecular fingerprints, scaffolds, molecular properties, and the number of NPs[58]. The multiple-variable plot was generated by GraphPad Prism software version 9.4.0, whereas the y-axis represents the area under the cyclic system recovery curve[59], the x-axis, represents the median of the fingerprint-based diversity computed with Molecular Access System (MACCS) keys (166-bits) and the Tanimoto coefficient[60], the bubble color represents the molecular properties of pharmaceutical interest, and the bubble size represents the number of NPs for each database.

**Drug-likeness.** The Osiris DataWarrior v05.02.01 software[61] was employed to calculate the drug-likeness score of the compounds from the PeruNPDB; the calculation is based on a library of 5300 substructure fragments and their associated drug-likeness scores. This library was prepared by fragmenting 3300 commercial drugs as well as 15,000 commercial non-drug-like Fluka NPs[61]. Frequency distribution of the obtained scores was performed at GraphPad Prism software version 9.4.0 for Windows, GraphPad Software, San Diego, California USA (http://www.graphpad.com), and plotted into stacked bar plots. Furthermore, the Lipinski Rule-of-5 (Ro5) is a set of four rules (logP, MW, and H-bond donor and acceptor cut-offs) for drug-likeness and oral bioavailability derived from a subset of 2245 drugs[62]. For this Lipinski's Ro5 KNIME node was employed to assess the number of violations to the rule for each compound on the PeruNPDB and plotted into pie charts. The US Food and Drug Administration (FDA)-approved drugs dataset[57], was employed as a reference, whereas the same procedures were applied to their compounds. Also, the chemical space representation was analyzed, and the procedures were the same as described earlier.

## Results
**PeruNPDB assemble.** In the present study, the assembly of the PeruNPDB, followed by its chemoinformatic characterization on molecular diversity and coverage of the chemical space was performed; to select the studies from which the NPs will further retrieve, a search with the MeSH Terms "*Peru*" AND "*Natural Products*" was performed in the Pubmed database, followed by the construction of a network map of the co-occurrence of MeSH terms. The workflow proposed in Fig. 1 was considered. The search resulted in 399 published papers between 1950-2021, whereas establishing the value of five as the minimum number of occurrences of keywords, a map with 194 keywords that reaches the threshold was constructed (Fig. 2A). In the analysis of the map, it is shown that six main clusters were formed, while terms such as "*Plant Extracts*", "*Plants, medicinal*", "*Phytotherapy*", "*Ethnopharmacology*", "*Ethnobotany*", "*Plants stems*", "*Plants bark*", and "*Seeds*", which are associated with NPs were observed in the first cluster (red color). Also, terms such as "*Peru*", "*Humans*", "*Animals*", and "*Male*", were recurrent terms. Although using the eligibility criterion established, 47 articles were selected which showed a 2000-2021-year range, and terms such as "*Flavonoids*", "*Sesquiterpenes*", and "*Anthocyanins*", were recurrent terms (Fig. 2B). Also, bibliographic data extracted from the selected articles analyzed: the "*Journal of Agricultural and Food Chemistry*", the "*Journal of Ethnopharmacology*", "*Phytochemistry*", and "*Planta Medica*" where the main peer-reviewed journals were the studies describing compounds extracted from Peruvian NPs were published (Fig. 2C). Furthermore, while retrieving the SMILES of the compounds from PubChem, DrugBank, and ChEMBL, it was observed that 242 structures were found in the repositories and that 38 needed to be generated in the ChemDraw tool. Ninety-five and five percent of the compounds were retrieved from plant or animal sources, respectively (Fig. 3A). The genus from which most of the compounds were extracted were Uncaria and Lepidium, with 11 and 10 percent, respectively (Fig. 3B). When analyzing the structure of the compounds with a classification system for small molecule structures, it is shown that 76 classes of NPs were found among the 280 NPs of the PeruNPDB, whereas anthocyanidins (N=25), aporphine alkaloids (N=11), cinnamic acids and derivatives (N=17), germacrane sesquiterpenoids (N=13), stigmastane steroids (N=10), and unsaturated fatty acids (N=22) were the most predicted classes of NPs (Fig. 4).
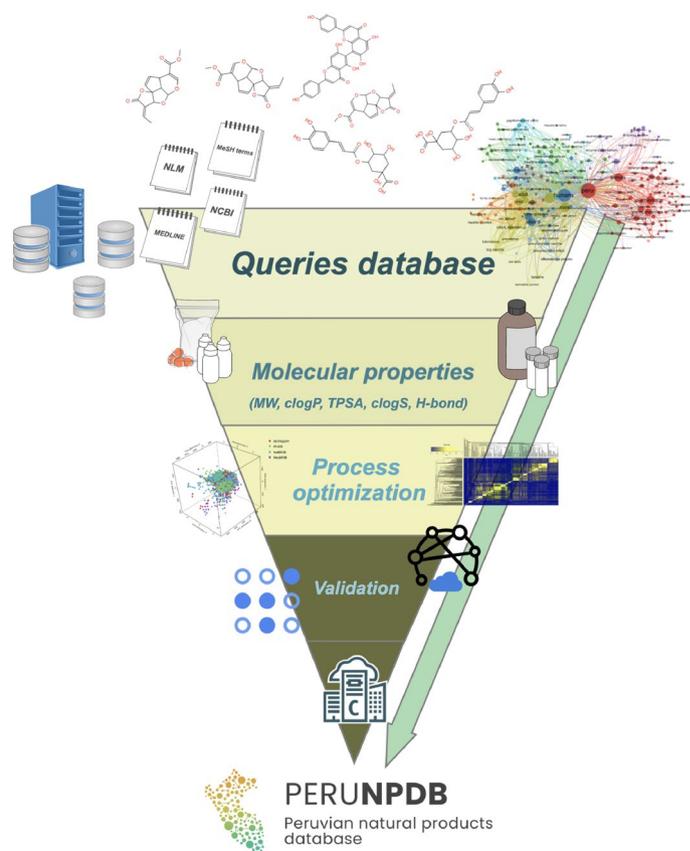
**Figure 1.** General workflow to generate and curate the first version of the Peruvian natural product database. The graph was edited in SmartDraw 2023 Software, LLC (Accessed April 14, 2023).

**Molecular properties.** Six physicochemical properties were calculated for all compounds in PeruNPDB and plotted into box plots, which include the distribution of the same properties of the three reference datasets, retrieved from the ZINC20 database (Fig. 5). To compare the results of the datasets, the coefficient of variation (CV) was calculated, which represents the ratio of the standard deviation to the mean and is considered a useful tool to statistically compare the degree of variation from one dataset to another[54]. Besides the results of the HBA, in which NuBBEDB obtain the highest CV (123.2%), the PeruNPDB showed the highest CV in MW, clogP, TPSA, clogS, and HBD with 46.58%, 84.49%, 112.8%, 50.08%, and 83.84%, respectively. Still, the results from TPSA, clogP, clogS, and HBD showed high statistical differences compared to AfroDB, BIOFAQUIM, and NuBBEDB, while showed no statistical difference in HBA results compared to the AfroNP database (Fig. 5).

**Visualization of the chemical space.** The chemical space visualization of PeruNPDB was conducted using PCA and t-SNE. Though the visual analysis of 3D-PCA shows that molecules in PeruNPDB share the chemical space roughly with NuBBEDB. Whereas in some regions the molecules of PeruNPDB are predominant (Fig. 6A). While the explained variance percentage of PC1, PC2, and PC3 was 50.24, 39.94, and 6.72, respectively. PeruNPDB, BIOFAQUIM, and NuBBEDB chemicals overlap in most of the chemical space represented, according to the 2D-t-SNE visual analysis (Fig. 6B).

**Diversity analysis.** The heatmap generated using the Tanimoto score matrix and the atom-pair-based fingerprints show that there is a similarity between the structures of the compounds of the PeruNPDB, AfroDB, BIOFAQUIM, and NuBBEDB (Fig. 6C). Also, a consensus diversity plot was used to evaluate the diversity of the PeruNPDB dataset, based on molecular fingerprints, scaffolds, and physicochemical properties. The Euclidean distance of the scaled properties was used to compute the property-based diversity of the PeruNPDB, AfroDB, BIOFAQUIM, and NuBBEDB databases. Data points on a continuous color scale are used to represent the values on the color CD plot. Darker colors signify less diversity, but brighter colors signify more diversity. Finally, different point sizes are used to illustrate how large or tiny the databases are, with smaller data points indicating databases with fewer molecules. The results showed that the diversity of compounds found in the PeruNPDB was the largest since it was found in the area where the highest diversity in scaffold and fingerprints should are found (Fig. 7), which is consistent with the results shown in the box plots (Fig. 6).
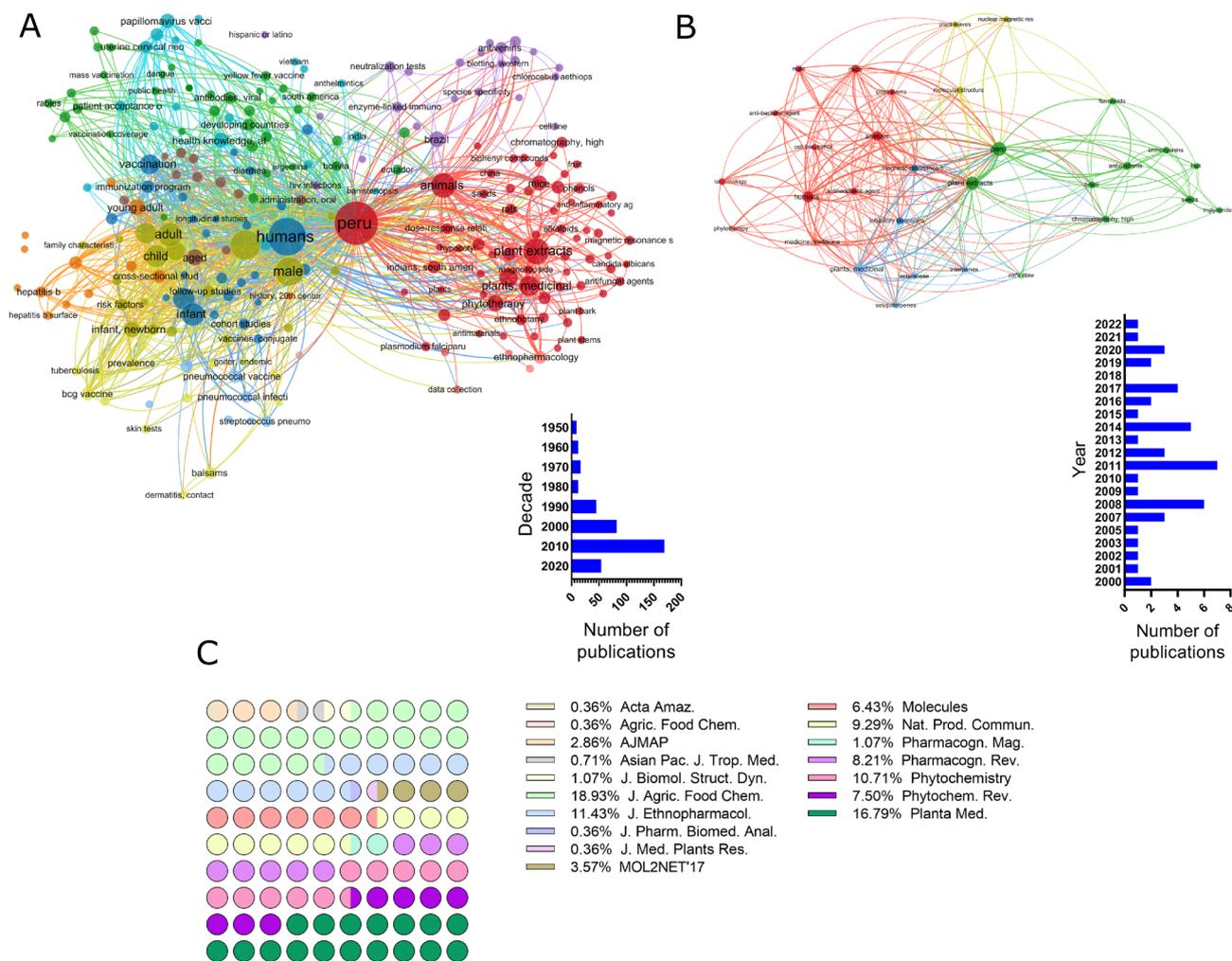
**Figure 2.** Bibliographic search for studies describing the characterization of Peruvian natural products. (**A**) Network map of the co-occurrence of MeSH terms. (**B**) Network map of articles selected from 2000 to 2021-year. (**C**) Bibliographic data extracted from selected articles that describe compounds extracted from Peruvian NPs.

**Drug-likeness.** Druglikeness assesses qualitatively the chance for a molecule to become an oral drug concerning bioavailability and is established from structural or physicochemical inspections of development compounds advanced enough to be considered oral drug candidates[63]. To assess the "*drug-like*" profile of the compounds from the PeruNPDB two approaches were performed; firstly, the frequency distribution of the drug-likeness score was analyzed, and the results showed that besides the differences in the number of compounds compared in both datasets a similar distribution among the compounds is observed (Fig. 8A). In the second approach, the number of violations to Lipinski's Ro5 was analyzed and the results showed that compounds with at least one violation represent the 85.82 and 76.35% of the FDA and PeruNPDB datasets, respectively (Fig. 8B). Also, the visual representation of the chemical space as PCAs (Fig. 8C) and t-SNE (Fig. 8D) indicates that some of the NPs are distributed in the same space as the already approved drugs. Whereas the explained variance percentage of PC1, PC2, and PC3 was 52.38, 37.64, and 5.54, respectively. The findings imply that because the compounds in PeruNPDB have chemical structures like those of approved medications, they can be used in virtual screening to find possible lead compounds or points for further optimization.

## Discussion

Peru has exceptionally high biodiversity, with numerous endemic species of mammals, reptiles, amphibians, flowering plants, and ferns, which is why has been described as a "*megadiverse*" country[64,65], but worldwide hotspot analysis for potential conflict between food security and biodiversity conservation points out Peru as a region that is especially at risk of biodiversity loss due to agricultural expansion[66]. Thus, the conservancy of biodiversity can be considered important since historically NPs have played a key role in drug discovery, especially for illnesses such as cancer, cardiovascular and infectious diseases[67], while the growing interest in NPs and their application is evidenced by a growth of the number of published databases of NPs, and collections of structures from various organisms, geographical locations, targeted diseases, and traditional applications[68]. Currently,
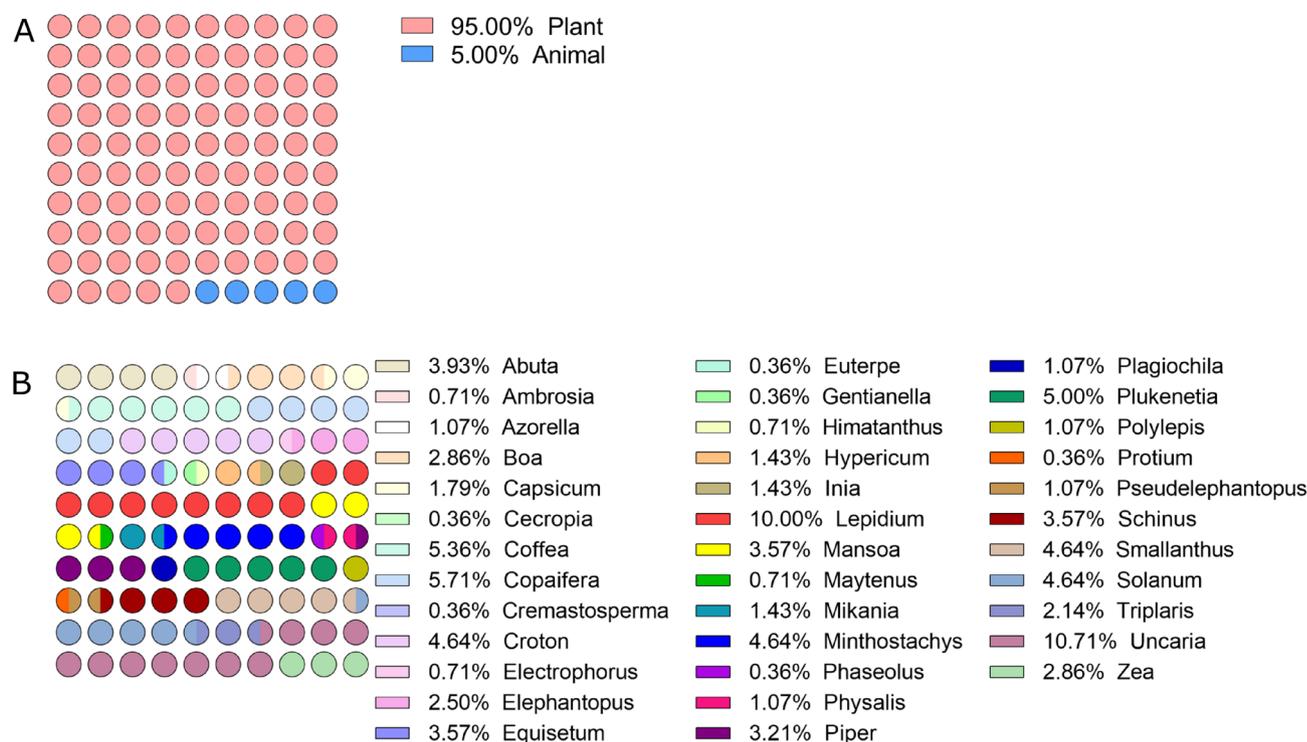
**Figure 3.** Dot plots showing the kingdom and genus of the species studied. (**A**) Compounds of Peruvian NPs found in PubChem, DrugBank, and ChEMBL databases. (**B**) Dot plot of the genus of the Peruvian NPs compounds obtained from the databases.

several NPs or NPs-derived molecules are employed in the treatment of distinct diseases, such as the antibiotic penicillin originally obtained from the fungi *Penicillium* spp.[69]; the analgesic aspirin, which is the most used drug in the world, derived from salicin extracted from the bark of the willow trees *Salix alba*[70]; and the immunosuppressant tacrolimus employed in the prevention of the rejection organ after transplants, obtained from bacteria *Streptomyces tsukubaensis*[71], are some examples. Besides, NPs and their derivatives have been considered promising options to improve treatment efficiency in cancer patients and decrease adverse reactions[72], whereas vinca alkaloids[73], taxane diterpenoids[74], camptothecin derivatives[75], and epipodophyllotoxin[76], are NPs-derived anticancer compounds clinically used as chemotherapeutics; while an example of the importance of biodiversity conservation is exemplified by the tree *Taxus brevifolia*, from which the chemotherapeutic drug paclitaxel was originally extracted, that was put on the list of endangered species[77,78]. According to the data, there are fewer compounds identified in the PeruNPDB than in AfroDB, BIOFAQUIM, and NuBBEDB, but the chemical diversity is also higher. Of the 280 compounds characterized, 95% came from plant sources, and 5% came from animal sources. But in the BIOFACQUIM and NuBBE databases as well as plant sources, compounds derived from fungi, propolis, bacteria, and marine organisms are also described. This partially explains the difference in the TPSA results of the PeruNPDB, since it has been reported that natural products from the animal kingdom have the highest TPSA due to the number of hydrogen bond donors and acceptors[79]. Furthermore, the Peruvian marine biodiversity hotspot located on the northern coast has been predicted to hold 501 species, 270 genera, and 193 families[80], as marine natural products have shown an interesting array of diverse and novel chemical structures with potent biological activities[81], which includes: Cephalosporin C an antibiotic derived from marine fungi *Cephalosporium*[82], Eribulin an anticancer drug derived from halichondrin B from the natural Japanese marine sponge *Halichondria okada*[83] and the antiviral, isolated from sponge *Tethya crypta*, nucleoside Ara-A[84]. Also, Peru is considered a diverse country that has a very broad microbial diversity richness, however, remains slightly studied and exploited[85,86]. Fungi, the eukaryotic microorganisms, produce a tremendous number of NPs with diverse chemical structures and biological activities[87], such as lovastatin, the first statin approved as a hyper-cholesterolemic medication by the FDA, most frequently produced by *Aspergillus terreus*[88], and cyclosporine A, a potent immunosuppressant that was initially used to prevent organ rejection, isolated from the fungal species *Tolypocladium inflatum* gams[89]. Besides that no current drug has been developed from propolis, it is considered a very rich and complex chemical composition, while about 300 different chemicals components isolated from it, and which composition fluctuates according to parameters such as plant source, seasons harvesting, geography, type of bee flora, climate changes, and honeybee species[90,91]; highlighting Artepillin C, extracted from Brazilian green propolis, that showed in vitro[92] and in vivo[93] anti-inflammatory potential. These emphasize the urgency to promote and enhance the study of Peruvian NPs quantitatively and qualitatively. Compounds from Peruvian medicinal plants have been evaluated for their antidiabetic[94], anticancer[95], antiviral[96], antibiotic[97], and antiparasitic activities[98]; however, most of the studies in the literature were in vitro performed over plants extracts, and little information about the potential of single compounds on these activities is described, while these promising
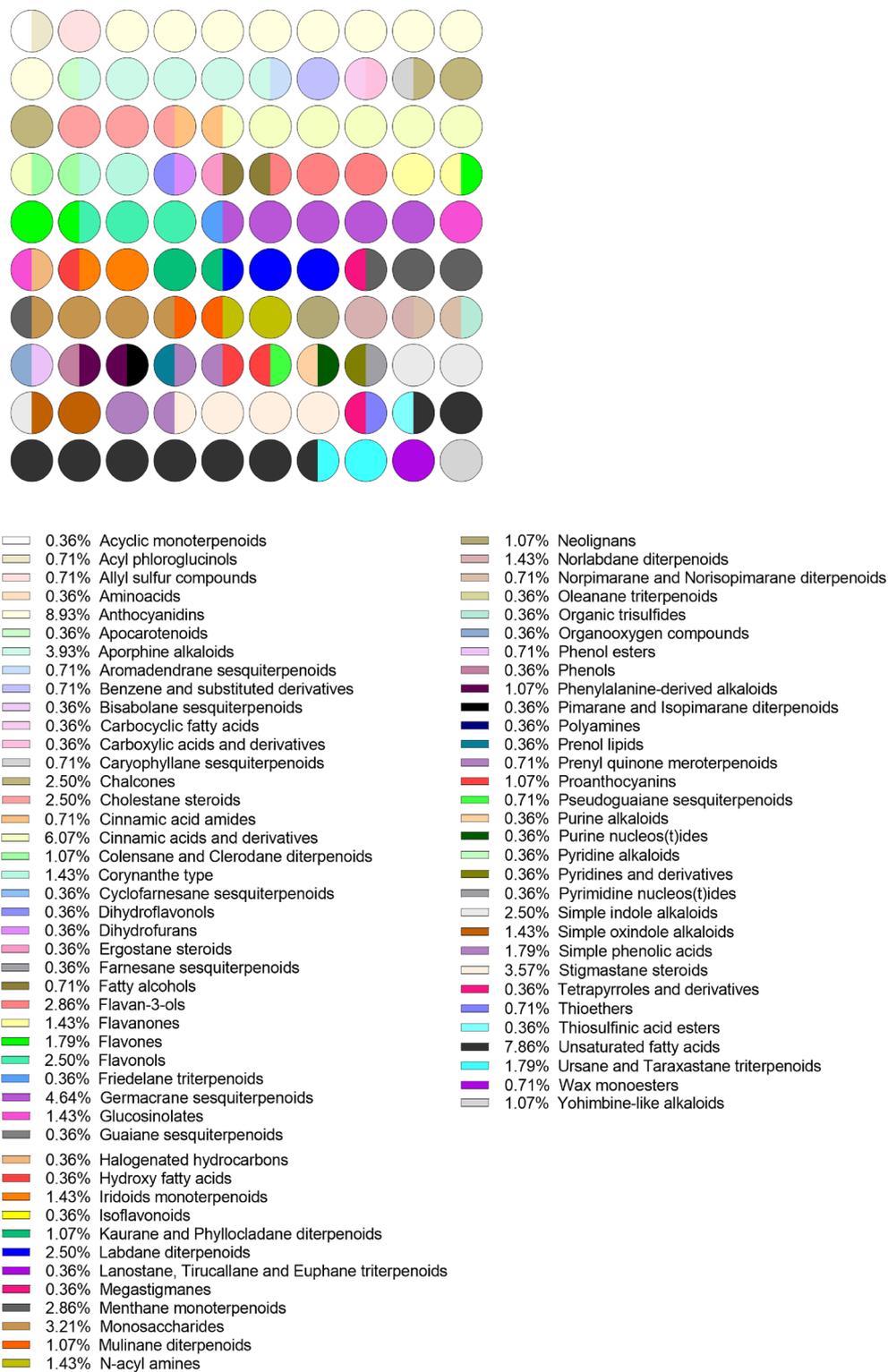
**Figure 4.** Dot plots showing the natural products classification.

| | | | |
|---|---|---|---|
| 0.36% | Acyclic monoterpenoids | 1.07% | Neolignans |
| 0.71% | Acyl phloroglucinols | 1.43% | Norlabdane diterpenoids |
| 0.71% | Allyl sulfur compounds | 0.71% | Norpimarane and Norisopimarane diterpenoids |
| 0.36% | Aminoacids | 0.36% | Oleanane triterpenoids |
| 8.93% | Anthocyanidins | 0.36% | Organic trisulfides |
| 0.36% | Apocarotenoids | 0.36% | Organooxygen compounds |
| 3.93% | Aporphine alkaloids | 0.71% | Phenol esters |
| 0.71% | Aromadendrane sesquiterpenoids | 0.36% | Phenols |
| 0.71% | Benzene and substituted derivatives | 1.07% | Phenylalanine-derived alkaloids |
| 0.36% | Bisabolane sesquiterpenoids | 0.36% | Pimarane and Isopimarane diterpenoids |
| 0.36% | Carbocyclic fatty acids | 0.36% | Polyamines |
| 0.36% | Carboxylic acids and derivatives | 0.36% | Prenol lipids |
| 0.71% | Caryophyllane sesquiterpenoids | 0.71% | Prenyl quinone meroterpenoids |
| 2.50% | Chalcones | 1.07% | Proanthocyanins |
| 2.50% | Cholestane steroids | 0.71% | Pseudoguaiane sesquiterpenoids |
| 0.71% | Cinnamic acid amides | 0.36% | Purine alkaloids |
| 6.07% | Cinnamic acids and derivatives | 0.36% | Purine nucleos(t)ides |
| 1.07% | Colensane and Clerodane diterpenoids | 0.36% | Pyridine alkaloids |
| 1.43% | Corynanthe type | 0.36% | Pyridines and derivatives |
| 0.36% | Cyclofarnesane sesquiterpenoids | 0.36% | Pyrimidine nucleos(t)ides |
| 0.36% | Dihydroflavonols | 2.50% | Simple indole alkaloids |
| 0.36% | Dihydrofurans | 1.43% | Simple oxindole alkaloids |
| 0.36% | Ergostane steroids | 1.79% | Simple phenolic acids |
| 0.36% | Farnesane sesquiterpenoids | 3.57% | Stigmastane steroids |
| 0.71% | Fatty alcohols | 0.36% | Tetrapyrroles and derivatives |
| 2.86% | Flavan-3-ols | 0.71% | Thioethers |
| 1.43% | Flavanones | 0.36% | Thiosulfinic acid esters |
| 1.79% | Flavones | 7.86% | Unsaturated fatty acids |
| 2.50% | Flavonols | 1.79% | Ursane and Taraxastane triterpenoids |
| 0.36% | Friedelane triterpenoids | 0.71% | Wax monoesters |
| 4.64% | Germacrane sesquiterpenoids | 1.07% | Yohimbine-like alkaloids |
| 1.43% | Glucosinolates | | |
| 0.36% | Guaiane sesquiterpenoids | | |
| 0.36% | Halogenated hydrocarbons | | |
| 0.36% | Hydroxy fatty acids | | |
| 1.43% | Iridoids monoterpenoids | | |
| 0.36% | Isoflavonoids | | |
| 1.07% | Kaurane and Phyllocladane diterpenoids | | |
| 2.50% | Labdane diterpenoids | | |
| 0.36% | Lanostane, Tirucallane and Euphane triterpenoids | | |
| 0.36% | Megastigmanes | | |
| 2.86% | Menthane monoterpenoids | | |
| 3.21% | Monosaccharides | | |
| 1.07% | Mulinane diterpenoids | | |
| 1.43% | N-acyl amines | | |

results can be explained by synergistic interaction or multi-factorial effects between compounds present in the plant extracts studied[99]. While pharmacodynamic synergy involves multiple substances acting on various receptor targets to enhance the overall therapeutic effect, and pharmacokinetic synergy involves substances with little to no activity helping the main active principle to reach the target by improving bioavailability or by reducing metabolism and excretion, this type of assay can hide the true potential of single molecules activity between different constituents of plant extracts[100]. Thus, the concerted effort of experimental NPs research with CADD is
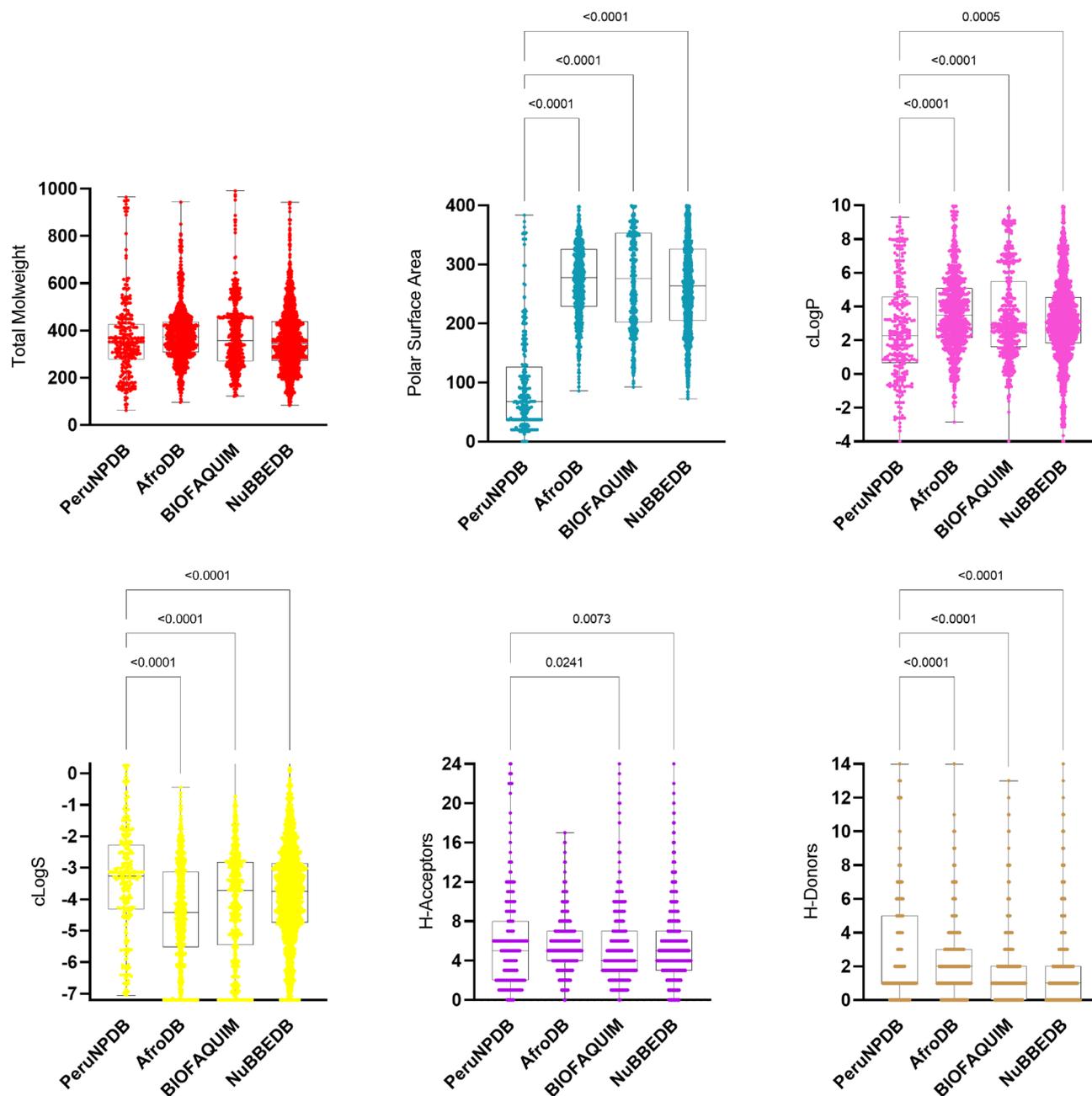
**Figure 5.** Box plots for the physicochemical properties of PeruNPDB and reference datasets.

continuously increasing; and recently, NPs from the Peruvian native plants *Smallanthus sonchilofolius*, *Lepidium meyenii* (40 compounds)[39], and *Uncaria tomentosa* (26 compounds)[101] were in silico analyzed for their antiviral activity against SARS-CoV-2. Also, the in silico polypharmaceutical potential of 84 NPs from *S. sonchifolius*, *L. meyenii*, *Croton lechleri*, *U. tomentosa*, *Minthostachys mollis*, and *Physalis peruvianus* was analyzed against Alzheimer's disease[102].

## Conclusion

Here we present the first version of PeruNPDB, a compound database of NPs from Peru that includes 280 compounds from plant and animal sources. PeruNPDB was constructed curated, and maintained by the Computational Biology and Chemistry Research Group from the Universidad Catolica de Santa Maria, and it is freely accessible through the website https://perunpdb.com.pe/. The PeruNPDB was envisioned as a tool for virtual screening, identifying promising compounds, serving as a springboard for further biotechnological products, and providing suggestions for conservation policies. The chemoinformatic characterization and analysis of the coverage and diversity of PeruNPDB in chemical space suggest broad coverage, overlapping with regions in the drug-like chemical space. The database contains an identification code (ID), the chemical name, bibliographic reference (name of the journal, year of publication, and DOI number), kingdom, genus, and species of the
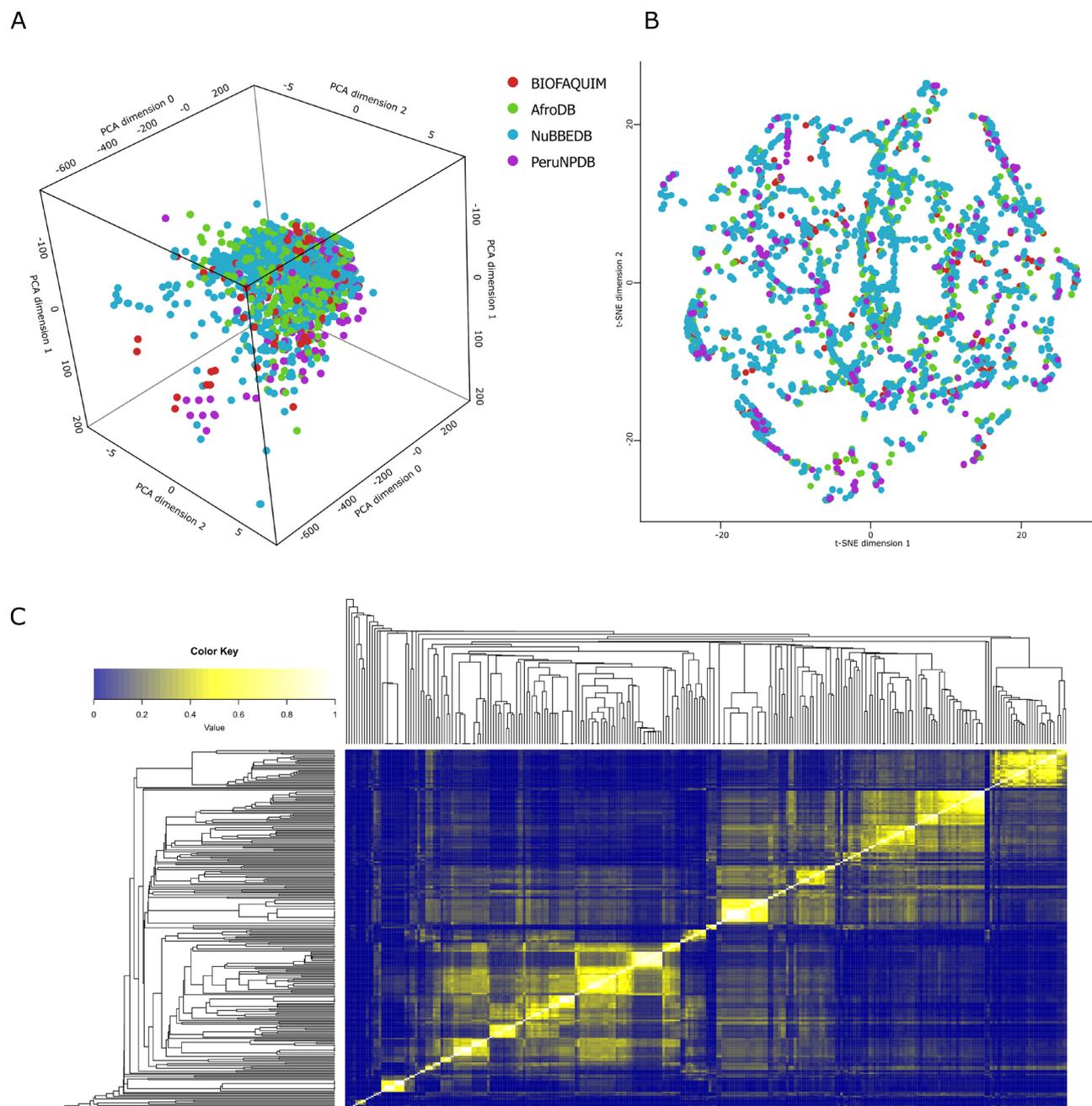
**Figure 6.** Visual representation of the chemical space of the PeruNPDB and reference datasets. (**A**) PeruNPDB 3D-PCA chemical space. (**B**) 2D-t-SNE visual analysis of the compounds PeruNPDB, AfroNP, BIOFAQUIM, and NuBBEDB. (**C**) Heatmap generated with Tanimoto scoring matrix of similar structures among compounds between PeruNPDB and control data sets.

natural product, SMILES notation, and classification of the natural product. In the future, we want to launch the PeruNPDB version 2 with new computed molecular descriptors, NP stereochemical data, and the possibility to download several structures at once. The web-based user interface will also be improved and kept, and new NPs from various taxonomic ranks that aren't included in the current edition will be added. Additionally, as we increase the quantity of NPs, we anticipate comparing the PeruNPDB with larger, more varied free datasets that are available in the literature. The complete PeruNPDB dataset for research purposes is available upon request and may be directed to and will be fulfilled by the lead contact Miguel Angel Chavez Fumagalli (mchavezf@ucsm.edu.pe).
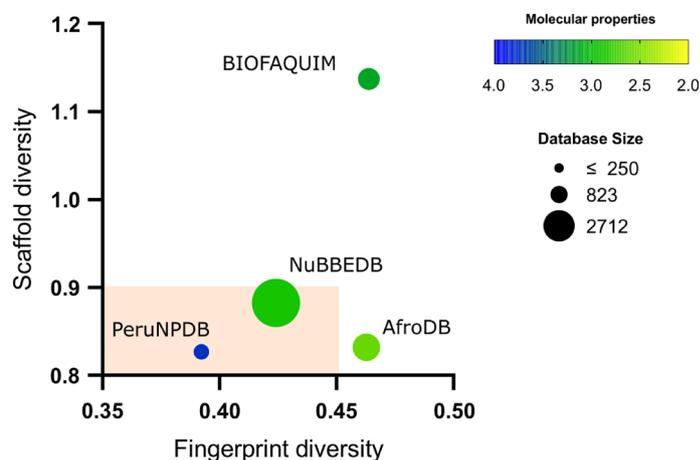
**Figure 7.** Consensus diversity plot comparing the global diversity of PeruNPDB with the reference data sets.
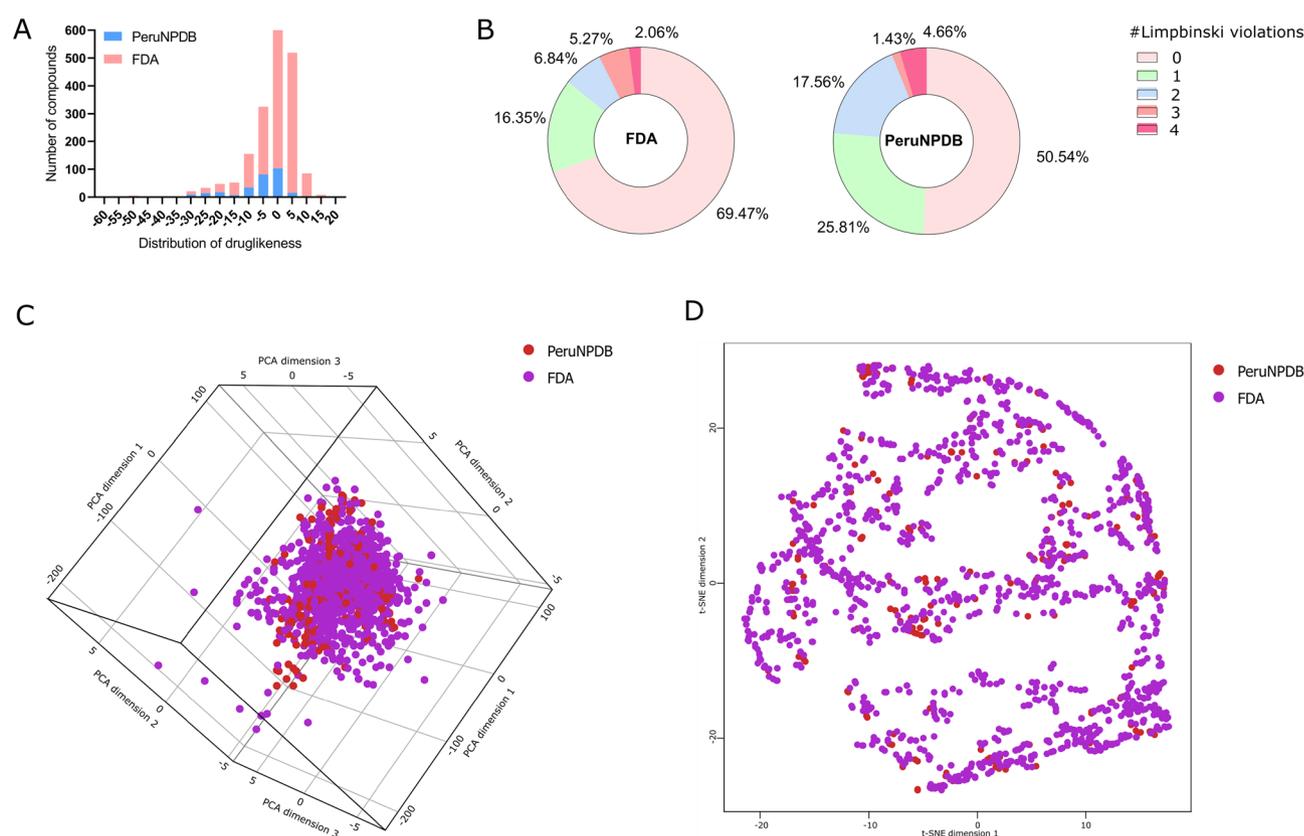


**Figure 8.** Druglikeness analysis of the PeruNPDB and the reference datasets. (**A**) The similar distribution between FDA compounds and PeruNPDB. (**B**) Lipinski's five rules from the FDA and PeruNPDB data sets. (**C**) Visual 3D representation of the chemical space as PCAs from the FDA and PeruNPDB data sets. (**D**) Representation t-SNE of FDA and PeruNPDB data sets.

## Data availibility

The datasets generated and/or analyzed during the current study are available in the PeruNPDB repository, https://perunpdb.com.pe/.

## References

1. Hanley, N. *et al.* The economic value of biodiversity. *Annu. Rev. Resour. Econ.* **11**, 355–375 (2019).

2. Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. & Worm, B. How many species are there on earth and in the ocean?. *PLoS Biol.* **9**, e1001127 (2011).
3. Sweetlove, L. Number of species on earth tagged at 8.7 million. *Nature* **23**, 1 (2011).
4. Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
5. Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M. & Gascon, C. Global biodiversity conservation: The critical role of hotspots. In *Biodiversity Hotspots*, 3–22 (Springer, 2011).
6. Fajardo, J., Lessmann, J., Bonaccorso, E., Devenish, C. & Munoz, J. Combined use of systematic conservation planning, species distribution modelling, and connectivity analysis reveals severe conservation gaps in a megadiverse country (peru). *PLoS ONE* **9**, e114367 (2014).
7. Shanee, S. *et al.* Protected area coverage of threatened vertebrates and ecoregions in Peru: Comparison of communal, private and state reserves. *J. Environ. Manag.* **202**, 12–20 (2017).
8. Institute, W. R. *Ecosystems and Human Well-being: Biodiversity Synthesis* (World Resources Institute, 2005).
9. Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).
10. Rai, M., Bhattarai, S. & Feitosa, C. M. *Ethnopharmacology of Wild Plants* (CRC Press, 2021).
11. Tresca, G., Marcus, O. & Politi, M. Evaluating herbal medicine preparation from a traditional perspective: Insights from an ethnopharmaceutical survey in the Peruvian Amazon. *Anthropol. Med.* **27**, 268–284 (2020).
12. Bussmann, R. W. & Sharon, D. Plantas medicinales de los andes y la amazonía-la flora mágica y medicinal del norte del perú. *Ethnobotany Res. Appl.* **15**, 1–293 (2016).
13. de Salud Convenio Hipólito Unanue, O. A. Plantas medicinales de la subregión andina (2014).
14. Achan, J. *et al.* Quinine, an old anti-malarial drug in a modern world: Role in the treatment of malaria. *Malar. J.* **10**, 1–12 (2011).
15. Calatayud, J. & González, Á. History of the development and evolution of local anesthesia since the coca leaf. *J. Am. Soc. Anesthesiol.* **98**, 1503–1508 (2003).
16. Schottenhammer, A. "Peruvian balsam": An example of transoceanic transfer of medicinal knowledge. *J. Ethnobiol. Ethnomed.* **16**, 1–20 (2020).
17. Lock, O., Perez, E., Villar, M., Flores, D. & Rojas, R. Bioactive compounds from plants used in Peruvian traditional medicine. *Nat. Prod. Commun.* **11**, 315–37 (2016).
18. Gonzales, G. F. & Valerio, L. G. Medicinal plants from Peru: A review of plants as potential agents against cancer. *Anti-Cancer Agents Med. Chem. (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)* **6**, 429–444 (2006).
19. Bussmann, R. W. The globalization of traditional medicine in northern perú: from shamanism to molecules. *Evid.-based Complement. Altern. Med.* **2013**, 291903 (2013).
20. Sabe, V. T. *et al.* Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review. *Eur. J. Med. Chem.* **224**, 113705 (2021).
21. Gimeno, A. *et al.* The light and dark sides of virtual screening: What is there to know?. *Int. J. Mol. Sci.* **20**, 1375 (2019).
22. Masic, I. & Ferhatovica, A. Review of most important biomedical databases for searching of biomedical scientific literature. *DSJUOG* **6**, 343–61 (2012).
23. Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A. & Steinbeck, C. Coconut online: Collection of open natural products database. *J. Cheminform.* **13**, 1–13 (2021).
24. Rutz, A. *et al.* The lotus initiative for open knowledge management in natural products research. *Elife* **11**, e70780 (2022).
25. Chen, C.Y.-C. TCM database@ Taiwan: The world's largest traditional Chinese medicine database for drug screening in silico. *PLoS ONE* **6**, e15939 (2011).
26. Mohanraj, K. *et al.* IMPPAT: A curated database of Indian medicinal plants, phytochemistry and therapeutics. *Sci. Rep.* **8**, 1–17 (2018).
27. Ntie-Kang, F. *et al.* AfroDb: A select highly potent and diverse natural product library from African medicinal plants. *PLoS ONE* **8**, e78085 (2013).
28. Pilon, A. C. *et al.* NuBBE$_{DB}$: An updated database to uncover chemical and biological information from Brazilian biodiversity. *Sci. Rep.* **7**, 1–12 (2017).
29. Scotti, M. T. *et al.* SistematX, an online web-based cheminformatics tool for data management of secondary metabolites. *Molecules* **23**, 103 (2018).
30. Pilón-Jiménez, B. A., Saldívar-González, F. I., Díaz-Eufracio, B. I. & Medina-Franco, J. L. BIOFACQUIM: A Mexican compound database of natural products. *Biomolecules* **9**, 31 (2019).
31. Gómez-García, A. & Medina-Franco, J. L. Progress and impact of Latin American natural product databases. *Biomolecules* **12**, 1202 (2022).
32. Acevedo, C. H., Scotti, L. & Scotti, M. T. In silico studies designed to select sesquiterpene lactones with potential antichagasic activity from an in-house asteraceae database. *ChemMedChem* **13**, 634–645 (2018).
33. do Carmo Santos, N., da Paixão, V. G. & da Rocha Pita, S. S. New *Trypanosoma cruzi* trypanothione reductase inhibitors identification using the virtual screening in database of nucleus bioassay, biosynthesis and ecophysiology (NuBBE). *Anti-Infect. Agents* **17**, 138–149 (2019).
34. Antunes, S. S., Rabelo, V.W.-H. & Romeiro, N. C. Natural products from Brazilian biodiversity identified as potential inhibitors of PknA and PknB of *M. tuberculosis* using molecular modeling tools. *Comput. Biol. Med.* **136**, 104694 (2021).
35. Herrera-Acevedo, C. *et al.* Selection of antileishmanial sesquiterpene lactones from SistematX database using a combined ligand-/structure-based virtual screening approach. *Mol. Divers.* **25**, 2411–2427 (2021).
36. Ccahuana, H. L. B. *et al.* In silico-based screening for natural products structural analogs as new drugs candidate against leishmaniasis. *bioRxiv* (2022).
37. Menezes, R. P. B. D., Viana, J. D. O., Muratov, E., Scotti, L. & Scotti, M. T. Computer-assisted discovery of alkaloids with schistosomicidal activity. *Curr. Issues Mol. Biol.* **44**, 383–408 (2022).
38. Rodrigues, G. *et al.* Ligand and structure-based virtual screening of lamiaceae diterpenes with potential activity against a novel coronavirus (2019-NCOV). *Curr. Top. Med. Chem.* **20**, 2126–2145 (2020).
39. Goyzueta-Mamani, L. D., Barazorda-Ccahuana, H. L., Mena-Ulecia, K. & Chávez-Fumagalli, M. A. Antiviral activity of metabolites from Peruvian plants against SARS-CoV-2: An in silico approach. *Molecules* https://doi.org/10.3390/molecules26133882 *(2021)*.
40. White, J. Pubmed 2.0. *Med. Ref. Serv. Q.* **39**, 382–387 (2020).
41. Rogers, F. B. Medical subject headings. *Bull. Med. Libr. Assoc.* **51**, 114–116 (1963).
42. Van Eck, N. J. & Waltman, L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics* **111**, 1053–1070 (2017).
43. Waltman, L. & Van Eck, N. J. A new methodology for constructing a publication-level classification system of science. *J. Am. Soc. Inf. Sci. Technol.* **63**, 2378–2392 (2012).
44. Weininger, D. Smiles, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988).
45. Kim, S. *et al.* Pubchem 2019 update: Improved access to chemical data. *Nucl. Acids Res.* **47**, D1102–D1109 (2019).

46. Wishart, D. S. *et al.* Drugbank 5.0: A major update to the DrugBank database for 2018. *Nucl. Acids Res.* **46**, D1074–D1082 (2018).
47. Mendez, D. *et al.* ChEMBL: Towards direct deposition of bioassay data. *Nucl. Acids Res.* **47**, D930–D940 (2019).
48. Evans, D. A. History of the Harvard ChemDraw project. *Angew. Chem. Int. Ed.* **53**, 11140–11145 (2014).
49. Sander, T., Freyss, J., von Korff, M. & Rufener, C. DataWarrior: An open-source program for chemistry aware data visualization and analysis. *J. Chem. Inf. Model.* **55**, 460–473 (2015).
50. Fillbrunn, A. *et al.* KNIME for reproducible cross-domain analysis of life science data. *J. Biotechnol.* **261**, 149–156 (2017).
51. Kim, H. W. *et al.* NPClassifier: A deep neural network-based structural classification tool for natural products. *J. Nat. Prod.* **84**, 2795–2807 (2021).
52. Djoumbou Feunang, Y. *et al.* ClassyFire: Automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* **8**, 1–20 (2016).
53. Bro, R. & Smilde, A. K. Principal component analysis. *Anal. Methods* **6**, 2812–2831 (2014).
54. Probst, D. & Reymond, J.-L. Visualization of very large high-dimensional data sets as minimum spanning trees. *J. Cheminform.* **12**, 1–13 (2020).
55. Cao, Y., Charisi, A., Cheng, L.-C., Jiang, T. & Girke, T. ChemmineR: A compound mining framework for R. *Bioinformatics* **24**, 1733–1734 (2008).
56. Team, R. C. *R: A Language and Environment for Statistical Computing.* http://www.R-project.org/ (R Foundation for Statistical Computing, 2013).
57. Irwin, J. J. *et al.* ZINC20—a free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* **60**, 6065–6073 (2020).
58. González-Medina, M., Prieto-Martínez, F. D., Owen, J. R. & Medina-Franco, J. L. Consensus diversity plots: A global diversity analysis of chemical libraries. *J. Cheminform.* **8**, 1–11 (2016).
59. Medina-Franco, J. L., Martínez-Mayorga, K., Bender, A. & Scior, T. Scaffold diversity analysis of compound data sets using an entropy-based measure. *QSAR Combinatorial Sci.* **28**, 1551–1560 (2009).
60. Kuwahara, H. & Gao, X. Analysis of the effects of related fingerprints on molecular similarity using an eigenvalue entropy approach. *J. Cheminform.* **13**, 1–12 (2021).
61. López-López, E., Naveja, J. J. & Medina-Franco, J. L. DataWarrior: An evaluation of the open-source drug discovery tool. *Expert Opin. Drug Discov.* **14**, 335–341 (2019).
62. Kralj, S., Jukič, M. & Bren, U. Comparative analyses of medicinal chemistry and cheminformatics filters with accessible implementation in Konstanz information miner (KNIME). *Int. J. Mol. Sci.* **23**, 5727 (2022).
63. Daina, A., Michielin, O. & Zoete, V. SwissABME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **7**, 1–13 (2017).
64. Rodríguez, L. O. & Young, K. R. Biological diversity of Peru: Determining priority areas for conservation. *AMBIO J. Hum. Environ.* **29**, 329–337 (2000).
65. McNeely, J. A. *et al. Conserving the World's Biological Diversity* (International Union for Conservation of Nature and Natural Resources, 1990).
66. Molotoks, A., Kuhnert, M., Dawson, T. P. & Smith, P. Global hotspots of conflict risk between food security and biodiversity conservation. *Land* **6**, 67 (2017).
67. Atanasov, A. G., Zotchev, S. B., Dirsch, V. M. & Supuran, C. T. Natural products in drug discovery: Advances and opportunities. *Nat. Rev. Drug Discov.* **20**, 200–216 (2021).
68. Sorokina, M. & Steinbeck, C. Review on natural products databases: Where to find data in 2020. *J. Cheminform.* **12**, 1–51 (2020).
69. Gaynes, R. The discovery of penicillin-new insights after more than 75 years of clinical use. *Emerg. Infect. Dis.* **23**, 849 (2017).
70. Montinari, M. R., Minelli, S. & De Caterina, R. The first 3500 years of aspirin history from its roots—A concise summary. *Vasc. Pharmacol.* **113**, 1–8 (2019).
71. Tanaka, H., Nakahara, K., Hatanaka, H., Inamura, N. & Kuroda, A. Discovery and development of a novel immunosuppressant, tacrolimus hydrate. *Yakugaku Zasshi J. Pharm. Soc. Jpn.* **117**, 542–554 (1997).
72. Choudhari, A. S., Mandave, P. C., Deshpande, M., Ranjekar, P. & Prakash, O. Phytochemicals in cancer treatment: From preclinical studies to clinical practice. *Front. Pharmacol.* **10**, 1614 (2020).
73. Martino, E. *et al.* Vinca alkaloids and analogues as anti-cancer agents: Looking back, peering ahead. *Bioorganic Med. Chem. Lett.* **28**, 2816–2826 (2018).
74. Oudard, S. *et al.* Cabazitaxel versus docetaxel as first-line therapy for patients with metastatic castration-resistant prostate cancer: A randomized phase iii trial-firstana. *J. Clin. Oncol.* **35**, 3189–3197 (2017).
75. Hertzberg, R. P., Caranfa, M. J. & Hecht, S. M. On the mechanism of topoisomerase i inhibition by camptothecin: Evidence for binding to an enzyme-DNA complex. *Biochemistry* **28**, 4629–4638 (1989).
76. Cao, B. *et al.* Cip-36, a novel topoisomerase II-targeting agent, induces the apoptosis of multidrug-resistant cancer cells in vitro. *Int. J. Mol. Med.* **35**, 771–776 (2015).
77. Mayor, S. Tree that provides paclitaxel is put on list of endangered species (2011).
78. Murage, P., Batalha, H. R., Lino, S. & Sterniczuk, K. From drug discovery to coronaviruses: Why restoring natural habitats is good for human health. *BMJ* **375**, n2329 (2021).
79. Pilkington, L. I. A chemometric analysis of deep-sea natural products. *Molecules* **24**, 3942 (2019).
80. Miloslavich, P. *et al.* Marine biodiversity in the Atlantic and Pacific coasts of South America: Knowledge and gaps. *PLoS ONE* **6**, e14631 (2011).
81. Jeong, G.-J., Khan, S., Tabassum, N., Khan, F. & Kim, Y.-M. Marine-bioinspired nanoparticles as potential drugs for multiple biological roles. *Mar. Drugs* **20**, 527 (2022).
82. Silber, J., Kramer, A., Labes, A. & Tasdemir, D. From discovery to production: Biotechnology of marine fungi for the production of new antibiotics. *Mar. Drugs* **14**, 137 (2016).
83. Swami, U., Shah, U. & Goel, S. Eribulin in cancer treatment. *Mar. Drugs* **13**, 5016–5058 (2015).
84. Sagar, S., Kaur, M. & Minneman, K. P. Antiviral lead compounds from marine sponges. *Mar. drugs* **8**, 2619–2638 (2010).
85. Vega, K. *et al.* Production of alkaline cellulase by fungi isolated from an undisturbed rain forest of peru. *Biotechnol. Res.Int.* **2012**, 934325 (2012).
86. Plata, E. R. & Lücking, R. High diversity of graphidaceae (lichenized ascomycota: Ostropales) in Amazonian Perú. *Fungal Divers.* **58**, 13–32 (2013).
87. Du, L. & Li, S. Compartmentalized biosynthesis of fungal natural products. *Curr. Opin. Biotechnol.* **69**, 128–135 (2021).
88. Al-Saman, M. A. *et al.* Optimization of lovastatin production by *Aspergillus terreus* ATCC 10020 using solid-state fermentation and its pharmacological applications. *Biocatal. Agric. Biotechnol.* **31**, 101906 (2021).
89. Colombo, D. & Ammirati, E. Cyclosporine in transplantation—A history of converging timelines. *J. Biol. Regul. Homeost. Agents* **25**, 493–504 (2011).
90. Hossain, R. *et al.* Propolis: An update on its chemistry and pharmacological applications. *Chin. Med.* **17**, 1–60 (2022).
91. Silva-Carvalho, R., Baltazar, F. & Almeida-Aguiar, C. Propolis: a complex natural product with a plethora of biological activities that can be explored for drug development. *Evidence-Based Complementary Altern. Medicine* **2015**, 206439 (2015).
92. Paulino, N. *et al.* Anti-inflammatory effects of a bioavailable compound, artepillin c, in Brazilian propolis. *Eur. J. Pharmacol.* **587**, 296–301 (2008).

93. Moura, S. A. L. D. *et al.* Aqueous extract of brazilian green propolis: primary components, evaluation of inflammation and wound healing by using subcutaneous implanted sponges. *Evidence-Based Complementary Altern. Medicine* **2011**, 748283 (2011).

94. Guillen Quispe, Y. N., Hwang, S. H., Wang, Z., Zuo, G. & Lim, S. S. Screening in vitro targets related to diabetes in herbal extracts from Peru: Identification of active compounds in *Hypericum laricifolium* Juss. by offline high-performance liquid chromatography. *Int. J. Mol. Sci.* **18**, 2512 (2017).

95. Bautista-Flores, A., Mayor, P. N. A. & Arellano, A. A. L. Cytotoxic effect of hydroalcoholic extract of annona muricata against a human cell line of gastric adenocarcinoma. *Vitae* **29**, 1–9 (2022).

96. Roumy, V. *et al.* Viral hepatitis in the Peruvian Amazon: Ethnomedical context and phytomedical resource. *J. Ethnopharmacol.* **255**, 112735 (2020).

97. Roumy, V. *et al.* In vitro antimicrobial activity of traditional plant used in mestizo shamanism from the Peruvian amazon in case of infectious diseases. *Pharmacogn. Mag.* **11**, S625 (2015).

98. Céline, V. *et al.* Medicinal plants from the Yanesha (Peru): Evaluation of the leishmanicidal and antimalarial activity of selected extracts. *J. Ethnopharmacol.* **123**, 413–422 (2009).

99. Williamson, E. M. Synergy and other interactions in phytomedicines. *Phytomedicine* **8**, 401–409 (2001).

100. Rasoanaivo, P., Wright, C. W., Willcox, M. L. & Gilbert, B. Whole plant extracts versus single compounds for the treatment of malaria: Synergy and positive interactions. *Malar. J.* **10**, 1–12 (2011).

101. Yepes-Pérez, A. F., Herrera-Calderon, O. & Quintero-Saumeth, J. *Uncaria tomentosa* (cat's claw): A promising herbal medicine against SARS-CoV-2/ACE-2 junction and SARS-CoV-2 spike protein based on molecular modeling. *J. Biomol. Struct. Dyn.* **40**, 2227–2243 (2022).

102. Goyzueta-Mamani, L. D. *et al.* In silico analysis of metabolites from Peruvian native plants as potential therapeutics against Alzheimer's disease. *Molecules* **27**, 918 (2022).

## Author contributions

Conceptualization: J.L.M.F. and M.A.C-F.; data curation: H.L.B-C., L.G.R., and M.A.C-F.; formal analysis: J.L.M.F. and M.A.C-F.; funding acquisition: G.D.D-C. and M.A.C-F.; investigation H.L.B-C., L.G.R., M.A.C-P., E.G.C-R., A.E.C-L.; methodology: J.L.M.F. and M.A.C-F.; writing—review and editing: H.L.B-C., L.G.R., G.D.D-C., J.L.M.F., and M.A.C-F. All authors have read and agreed to the published version of the manuscript.

## Funding

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to M.A.C.-F.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.