Redefining floristic zones on the Korean Peninsula using high-resolution georeferenced specimen data and self-organizing maps (SOMs)

Song Hie Jung[1] and Yong-chan Cho[1]

[1]Korea National Arboretum

June 24, 2020

Abstract

The use of biota to analyze the spatial range and distribution of biogeographic regions is essential to gain a better understanding of the ecological processes that cause biotic differentiation and biodiversity at multiple spatiotemporal scales. Recently, the collection of high-resolution biological distribution data (e.g., specimens) and advances in analytical theory have led to their quantitative analysis and more refined spatial delineation. This study was conducted to redefine floristic zones in the southern part of the Korean Peninsula and to better understand the eco-evolutionary significance of the spatial distribution patterns. Based on the distribution data of 309,333 vascular plant species in the Korean Peninsula, we derived floristic zones using self-organizing maps. We compared the characteristics of the derived regions with those of historical floristic zones and ecologically important environmental factors (climate, geology, and geography). In a clustering analysis of the floristic assemblages, four distinct regions were identified, namely, the cold floristic zone (Zone I) in high-altitude regions at the center of the Korean Peninsula, cool floristic zone (Zone II) in high-altitude regions in the south of the Korean Peninsula, warm floristic zone (Zone III) in low-altitude regions in the central and southern parts of the Korean Peninsula, and maritime warm floristic zone (Zone IV) including the volcanic islands of Jejudo and Ulleungdo. A total of 1,099 taxa were common to the four floristic zones. Zone IV had the highest abundance of specific plants (those found in only one zone), with 404 taxa. This study improves floristic zone definitions using high-resolution regional biological distribution data. It will help better understand and re-establish regional species diversity. In addition, our study provides key data for hotspot analysis techniques required for the conservation of plant diversity.

Introduction

Using biota to investigate biogeographic regions, including their spatial range and distribution patterns, is key to providing a better understanding of the ecological processes that create biotic differentiation and biodiversity at multiple spatiotemporal scales (Ricklefs, 2004). In particular, as an important step in understanding the spatial structure of biodiversity, delineating the spatial range of these regions has been a foundational part of basic and applied research in biogeography, ecology, earth science, and conservation ecology (Brum et al., 2017; Graham and Hijmans, 2006; Ibanez-Erquiaga et al., 2018; Kreft and Jetz, 2010; Lenormand et al., 2019; Olson et al., 2001; Ricklefs, 2004; Sun et al., 2008). Until the late 20[th]century, the proposed biological regions were based on limited data, convenience, and the opinions of experts (White, 1983). Without distinct criteria, floristic zones were also suggested based on delineating the distribution of endemic plants (Takhtajan, 1986), or accumulating floristic checklist data and applying spatial statistics (McLaughlin, 1989). Relatively recently, there have been many proposals for defining geographical spaces based on plants (Gonzalez-Orozco et al., 2014; Kreft and Jetz, 2010; Lenormand et al., 2019; Vilhena and

1

Antonelli, 2015). Difficulties exist with regard to producing precise mapping and understanding the regional patterns of biodiversity because of issues surrounding the fidelity and reliability of floristic surveys, and because flora is closely related to environmental gradients (development and climate) and their complexity (Gonzalez-Orozco et al., 2014). If these challenges can be overcome, the delineation of biological spaces based on high-fidelity, reliable data will provide new metrics and perspectives through biogeographic regionalization (Lenormand et al., 2019).

The first attempt at biogeographic regionalization of the Korean Peninsula, made around 100 years ago, proposed convenient northern, central, and southern divisions based on descriptions of plant and vegetation characteristics (Nakai, 1919). Subsequently, more finely differentiated floristic zones were defined by redistributing the divisions of Nakai (1919) (Lee and Yim, 1978). Other authors proposed pseudo-floristic zones from the perspective of vegetation and climate (Yim and Kira, 1975). All these previous floristic zones for the Korean Peninsula (including pseudo-floristic zones), which were mostly developed from expert opinion or for convenience, have been structured around homogeneous bands based on the relationship between latitude and mean annual air temperature. In the neighboring country China, large-scale banded or planar vegetation-climate zones have been delineated using a wetness index (Sun et al., 2008). However, regions with a complex mountainous structure show marked changes in elevation and topography over short distances and, when this is combined with human influence, it results in especially complicated spatiotemporally-driven biogeographic regions (see the map of floristic zones on the Korean Peninsula in Appendix A1) (Lenormand et al., 2019).

Given the lack of regionalization based on biological distribution data (e.g., specimens), pseudo-biogeographical divisions have also been developed for conservation at large spatial scales (Olson et al., 2001; Sun et al., 2008). When delineating biogeographical zones, it is necessary to maximize differences between zones while also maximizing the homogeneity of the taxonomic assemblages within them (Stoddart, 1992). Improving the analytical accuracy through the quantitative accumulation of organism distribution data, informatization of geography, and large-scale distribution data—which has previously been a challenge for the extraction of floristic and other biological zones—enables quantitative and rigorous regionalization (Linder et al., 2012). In recent studies, floristic zones have been delineated using accumulated data to incorporate information about plant distribution (Gonzalez-Orozco et al., 2014). The reliability of point data for the distribution of organisms can only be ensured by using specimen data. Because of the use of global positioning systems (GPS), plant specimens that include spatial information are being collected. From the late 20th century, accurate and extensive plant catalogs and data on the distribution of specimens have been collected (e.g., Korea National Arboretum, 2016), and floristic regions are being defined at the regional and national level using plant location data (Gonzalez-Orozco et al., 2014; Korea National Arboretum, 2016; Lenormand et al., 2019). It is, therefore, possible to develop approaches to delineate floristic zones using these data, reliably and accurately. In particular, in the southern part of the Korean Peninsula, the plant specimens that have been collected and the distribution maps that have been composed since 2000 can be evaluated with reliable large-scale data.

To analyze accumulated species distribution data, artificial neural networks (ANNs) are increasingly used as an alternative to traditional statistics for the analysis of multi-dimensional data (Chon, 2011; Cottrell et al., 2018; Snedden, 2019). Specifically, self-organizing maps (SOMs), an ANN-based technique using unsupervised learning, are suggested as an alternative to conventional primary component analysis (Ahn et al., 2018; Chon, 2011). Essentially, SOMs are classified as a non-linear sequence analysis method, since the training data set is non-linearly projected onto a lesser dimensional space (generally two dimensions) approximating the probability density function (Kohonen, 1995; Snedden, 2019). Unlike statistical approaches using mediator variables, SOMs do not make assumptions related to the correlations between variables or the distribution of variables (Chon, 2011; Giraudel and Lek, 2001; Snedden, 2019), and so are suitable for use with species presence or absence data (Céréghino et al., 2005). Efforts to derive and visualize zones from plant species distribution data using conventional univariate statistical analyses have generally assumed the response of species data to environmental gradients, in accordance with Eigen-based analytical approaches, but these analyses are limited in cases when the shape of the species-abundance response (e.g., linear or

unimodal) is not clear (Ahn et al., 2018; Liu et al., 2006; Snedden, 2019). Recently, numerous collections of distribution point data have been used for biological regionalization, but it is not possible to ascertain the relationships between species as variables. SOMs reduce multidimensional data to two or three dimensions, making them useful for typification analyses using distribution points for a large number of plant species (including regionalization).

Quaternary glacial–interglacial oscillations have been an important mechanism in shaping the current distribution of plants (Ricklefs, 1987). The Korean Peninsula, which is topographically composed of around 80% mountains, is characterized by a backbone mountain range (the Baekdudaegan Mountains) running north to south with sub-ranges branching off. The major mountains ([?]1000 m above sea level) are considered to be a single glacial refugium based on altitude rather than latitude, and they have a mixture of boreal and temperate flora (Chung et al., 2017b; Chung et al., 2018; Kim et al., 2014). The botanical importance of peninsulas and mountainous regions is well established because of their topographic characteristics (Medail and Diadema, 2009). The mountains that form the core topography of the Korean Peninsula possess a floristic composition that has been affected by latitude and spatiotemporal gradients, and this might be significant for its mutually distinct functions and evolutionary spaces. The flora of the Korean Peninsula is fundamentally controlled by a mixture of boreal and temperate abiotic conditions, and is affected, like other regions, by the agricultural and urbanization activities of humans.

The accumulation of a large volume of recent distribution point data and the application of analytical methods suited to the nature of the data have resolved the previous difficulties of floristic regionalization, making it possible to propose rigorous floristic zones according to the actual distribution of species. To date, however, it has been difficult to find case studies of accumulated, high-resolution, georeferenced specimen data for plants, or to find studies related to geographic regionalization using a SOM. The present study was conducted with the purpose of redefining the floristic zones in the southern part of the Korean Peninsula with SOMs and understanding the eco-evolutionary significance of the spatial distribution patterns. We used point distribution data for vascular plants collected at high resolution in the southern part of the Korean Peninsula between 2003 and 2015. We aimed to (1) derive floristic delimitations, (2) identify the correlations with ecologically important environmental factors, and (3) discuss the eco-evolutionary significance of the derived regions for floristic assemblages.

Materials and methods

Study region

Our study was conducted in the southern part of the Korean Peninsula, which is located the East Asia region (33deg–38deg N, 125deg–131deg E; Fig. 1). The total area of the study region is 100,033 km², and the human population is 51 million (Ministry of Land Infrastructure and Transport, 2016). The mean annual air temperature range ranges from 10–15 degC, and the range of the mean annual precipitation is 1000–1900 mm (Korean Meteorological Administration, 2020).

The Korean Peninsula is situated adjacent to the west Pacific Ocean and is surrounded by water in three directions—east, west, and south. It is a temperate region with four distinct seasons associated with the East Asian monsoon that occurs on the far eastern side of the Asian continent (Yi, 2011). Winter (December–February) is cold and dry because of the formation of the strong Siberian anticyclone from the Tibetan Plateau, while summer (June–August) is hot and humid, with around 70% of the annual precipitation focused in this period (Korean Meteorological Administration, 2020) (Appendix A2). Although the Korean Peninsula is located at the eastern edge of the Eurasian continent, the humid air supplied from the Yellow Sea to the west affects the diversity and distribution of local plants.

The Korean Peninsula comprises a large number of mountains centered in the Baekdudaegan Mountain Range, and only 22.5% of the peninsula is flat land (Appendix A3) (Ministry of Land Infrastructure and

3

Transport, 2016). In addition, although the elevation is not generally high, the region displays complex tectonic characteristics with a relatively diverse topography. Because the altitude gradients are shallower than those in other regions in East Asia, the borders between mountains and plateaus are relatively indistinct, making the region well suited for the spatiotemporal movement of plants (Ministry of Land Infrastructure and Transport, 2016).

There are around 4,300 known species of vascular plant on the Korean Peninsula (with approximately 3,000 species in the southern part), including 280 species of pteridophyte, 53 species of gymnosperm, and 3,963 species of angiosperm. In terms of specialized genera, *Pentactina* , *Echiosophora* , *Abeliophyllum* ,*Hanabusaya* , *Mankyua* , and *Megaleranthis* are present. According to the Whittaker biome classification (Whittaker, 1962), the southern part of the Korean Peninsula is mostly occupied by temperate seasonal forest biomes but may also contain some temperate rain forest and woodland/shrubland biomes (Fig. 1). In terms of the remnant vegetation landscape of the Korean Peninsula, strong policies to promote agriculture throughout the Joseon Period (1392–1910) led to a large decrease in forests and an increase in grassland and shrubland habitats. Later, in the southern part of the Korean Peninsula, the South Korean government pursued policies to promote forests from the 1970s, resulting in most natural habitats being located in forests (Cho et al., 2018). Currently, approximately 30.3% of the southern part of the Korean Peninsula is urbanized or used for agriculture and 63.8% is occupied by forests, with other land covers accounting for the remaining 5.9% (Ministry of Land Infrastructure and Transport, 2016).

Plant distribution data

We used vascular plant distribution data based on specimen and coordinate data for plants collected between 2003 and 2015 in the southern part of the Korean Peninsula (Korea National Arboretum, 2016). The vascular plant distribution maps contained coordinate data for 309,333 specimens, corresponding to 2,954 taxa in 175 families and 919 genera. For analysis, a grid system was overlaid on a national topographic map to combine the taxonomic groups located in each cell of the grid (cell size, 11.2 km x 13.9 km) with the location coordinates in a single data set (Graham and Hijmans, 2006; Lenormand et al., 2019). All 771 grid cells were used in the analysis, but some large urban regions were excluded from the floristic survey conducted by the Korea National Arboretum, and so these were left as empty cells.

Analysis of floristic assemblage clusters and characteristics

Using distribution data for the 771 grid cells and 2,965 plant taxa, a SOM training data set was constructed in the form of a presence-absence matrix (771 rows x 2,954 columns) (Fig. 2). The 'kohonen' R package was used for the SOM algorithm (Wehrens and Kruisselbrink, 2018), and the output layer was composed of 81 output nodes arranged in a square lattice. To determine the types, hierarchical cluster analysis was applied to the weight vectors of the SOM map units after conversion to Euclidean distance metrics (via the function hclust in R using the complete linkage method). The optimal number of types was calculated by applying the silhouette coefficient to the range of 2–15 types (Rousseeuw, 1987). In mapping the regionalization results, the grid cells that were empty because of exclusion from the survey were filled using the maximum frequency value from the surrounding eight cells. Since some island regions (Ulleungdo and Dokdo) showed heterogeneous values because of their distance from the adjacent grid cells, mapping was performed using type values within the local range.

The correlations in species composition between the floristic zones were analyzed using Venn diagrams (in the 'VennDiagram' package) based on lists of species in each zone (Chen and Boutros, 2011). After producing species catalogs for each zone, the common taxa (those appearing in all zones) and specific taxa (those appearing in only specific zones) were distinguished. Then, floristic compositions were investigated by analyzing the identification of specific taxa at the family level.

4

Environmental data and analysis

Geographic and climate factors were analyzed as macro-environmental factors, using the defined floristic zones. For geographic factors, the latitude and longitude were used, and for climate factors, air temperature and precipitation data—provided by the Korean Meteorological Administration (2020) and collected from 583 points between 1970 and 2010—were used. In addition to the direct environmental data, the warmth index (WI) and coldness index (CI) were calculated and used as indirect climate data (Kira, 1945) (Eq. 1 and 2). The values for these environmental factors were converted to values covering the whole southern part of the Korean Peninsula by linear interpolation, accounting for topography and altitude, with ArcGIS program (ver. 10.0). The mean values for the environmental factors in each grid cell were then calculated and used in the analysis.

$$\text{WI} = \sum_{1}^{n} (t - 5) : t > 5$$

$$\text{CI} = -\sum_{1}^{n} (t - 5) : t < 5 \backslash n$$

As physical factors affecting plant distribution, parent materials, topography, effective soil depth, and soil texture for the southern part of the Korean Peninsula were used (Rural Development Administration, 2010). Parent materials were categorized as acidic rock, metamorphic rock, sedimentary rock, quaternary deposit, volcanic ash, and other; topography was categorized into mountain, hill, pediment, interrill area, fan, lava terrace, or other; effective soil depth was categorized into four classes (<20 cm, 20–50 cm, 50–100 cm, and >100 cm), and soil texture was categorized as sandy gravel, silt and sandy loam, clay loam, and clay (Appendix A4).

To test the effect of environmental factors on the floristic composition and zonation, the geographic and climate (mean annual temperature, annual precipitation, warmth index, and coldness index) data were analyzed using the one-way analysis of variance (ANOVA) and Tukey's tests (Zar, 1984). The categorical physical factors (parent materials, topography, effective soil depth, and soil texture) were analyzed using box plots for each zone. The "ggplot2" R package was used for data visualization (Wickham, 2016). Statistical analyses were performed using R (R Core Team, 2019).

Results

Flora extraction and taxon distribution

After the coordinates of the plant specimens were converted to floristic composition data for the 771 grid cells, the phytogeographic structure of the southern part of the Korean Peninsula was divided into four zones (maximum silhouette coefficient: 0.7968, Table 1, Figs. 3 and 4). The peninsula was divided into three inland zones—the cold floristic zone (hereafter, Zone I), corresponding to the high-altitude regions in the central part of the peninsula; the cool floristic zone (hereafter, Zone II), corresponding to the high-altitude regions in the southern part of the peninsula; and the warm floristic zone (hereafter, Zone III), corresponding to the lowlands in the central and southern parts of the peninsula—and the maritime warm floristic zone (hereafter, Zone IV), including Jejudo and Ulleungdo.

Although these zones formed large patches and occupied large areas, smaller patches were scattered within the other floristic zones (Fig. 4). This is because Zone III, which was connected to Zone I, contained most of the major cities in a given region or broadscale production (forestry and agriculture) regions. Our study region also included the military demarcation line between North and South Korea. The region near the

5

military demarcation line (near the 38th parallel, length, 248 km; area, 907 km^2) is characterized by a high level of disruption and environmental management activities.

Zone III—broadly covering the lowlands of the southern part of the Korean Peninsula—occupied the largest area (72.0%), followed by Zones I (12.3%) and II (11.5%) in regions with high-altitude mountains, and Zone IV (4.2%), which included coastal and island regions (Table 1, Fig. 4). Unlike the area gradient, the order of species abundance (from highest to lowest) was: Zone III (2,379 taxa), Zone IV (2,200 taxa), Zone I (1,700 taxa), and Zone II (1,668 taxa). The species abundance per classified grid cell was highest in Zone I (464.7 ± 9.9 taxa), followed by Zone II (331.7 ± 11.5 taxa), Zone IV (298.6 ± 14.0 taxa), and finally Zone III (79.3 ± 3.4 taxa), which contained a large number of developed regions.

Geographic range and climatic environment in each floristic zone

All the floristic zones showed significant differences in the mean longitude and latitude, mean annual temperature, annual precipitation, WI, and CI per grid cell ($p < 0.001$, Fig. 5a–f). While Zone I was focused in the central part of the Korean Peninsula, Zone II was mostly situated in the center of the southern part of the peninsula, although there were scattered small patches within the central region adjacent to Zone I or inside Zone III (Figs. 4 and 5a–b). Zone IV was mostly located on the coast and islands in the west and south of the peninsula, as well as some inland areas in the west. Zone III showed a relatively broad longitudinal and latitudinal range, since it was distributed throughout the study region across the whole southern part of the Korean Peninsula.

When the climatic environments were compared between the floristic zones, differences were observed in mean annual air temperature, mean precipitation, WI, and CI (Fig. 5c–f). Mean annual air temperature showed a gradually increasing trend from Zones I to IV. Annual precipitation showed the opposite trend, but there was no clear difference between Zones I and II. As indirect climate factors, mean WI and CI also showed clear differences between zones. In accordance with its wide geographic distribution, Zone III showed the largest ranges for these climate factors.

The physical environment in each floristic zone

The parent materials in the southern part of the Korean Peninsula are predominantly acidic rocks (48.5%) and metamorphic rocks (26.3%), and the ratio of plateaus is very low. The most common soil depth is 50–100 cm (39.1%), and 93.7% of the soil consists of silt and sandy loam (51.6%) or clay loam (42.1%) (Ministry of Land Infrastructure and Transport, 2016) (Appendix A4).

The parent materials across the four floristic zones consisted of over 70% acidic and metamorphic rocks; however, Zone IV, which includes the volcanic island of Jejudo and the oceanic island of Ulleungdo, showed an especially high ratio of volcanic ash (23.7%) among its parent materials (Table 2). In addition, Zone IV showed a far higher proportion of lava terrace as its terrain type and clay loam for its soil type, and the soil depth classes were more evenly distributed compared to the other zones (Fig. 6).

The relationships between species composition in each floristic zone

A total of 1,099 common taxa was found in all zones (Fig. 7). Interestingly, the zone with the most specific taxa was Zone IV (404 specific taxa, 18.4%), which had the smallest area. This was followed by Zone III (192, 8.1%), Zone I (72, 4.2%), and Zone II (25, 1.5%).

When specific taxa were analyzed at the family level (Fig. 8), Zone III—generally consisting of low-lying hilly terrain heavily affected by human activity and development—showed higher diversity of Poaceae (18 out of 192 taxa) compared to Zones I and II, while the marine and coastal Zone IV was characterized by a high diversity of Orchidaceae (27 out of 404 taxa), Asteraceae, and Rosaceae. Among the common species

6

appearing in all zones, the families with the highest diversity, in descending order, were Asteraceae (113 taxa), Poaceae (98 taxa), Cyperaceae (68 taxa), Fabaceae (55 taxa), and Rosaceae (47 taxa) (Fig. 8).

Discussion

Historically, biogeographic spaces have predominantly been delineated using the distribution of organisms according to specific knowledge (e.g., partially usable data, such as endemism or the distribution of evergreen trees), or corresponding geographical and climatic factors, rather than using actual recorded biota (Takhtajan, 1986; White, 1983). Historical definitions of vegetation-climate and floristic zones on the Korean Peninsula have used these methods (Lee and Yim, 1978; Yim and Kira, 1975). Recently, there have been attempts to review bioregions using quantitative analytical techniques, with some showing similar, and some different, results to the historical regions (Gonzalez-Orozco et al., 2014; Lenormand et al., 2019). The present study differs fundamentally from previous approaches to phytogeographic regionalization because we used high-resolution, georeferenced specimen data for the southern part of the Korean Peninsula. Although restricted to the southern regions, our results provide clearly improved phytogeographic zones through our use of SOMs and georeferenced data for recently collected plant specimens. The spatial distribution of floristic zones and the basic taxonomic composition in the southern part of the Korean Peninsula reflect a combination of repeated migrations, expansions, and reductions of species associated with climatic variation, physical and geographic factors, and human activity patterns, including the political situation (Valladares et al., 2015). In addition, we reconfirmed differences identified in historical studies in geographic and climatic conditions, which are deterministic factors in the spatial patterns of floristic assemblages (Lee and Yim, 1978; Yim and Kira, 1975). We were also able to verify the correlations between zones and the distribution of zones based on the characteristics of the floristic assemblages and the physical conditions where they are found, such as the topographic characteristics of the Korean Peninsula. In particular, we verified the importance of the topography of the Korean Peninsula (especially the connected mountainous regions and isolated mountains), which has been emphasized in recent studies (Chung et al., 2017b), and its contribution to shifts in distribution related to climate change and the process of floristic reassembly. As well as the deterministic factors for floristic assembly that have been used historically (e.g., climatic factors), more diverse ecologically important environmental factors will need to be considered at the same spatial range as the southern part of the Korean Peninsula.

Spatial characteristics of the new floristic zones

The transitional changes in the flora of the southern Korean Peninsula (from warm to cool), and the effects of physical and human activity were investigated via the derived floristic zones. The historical floristic zones of the Korean Peninsula were constructed to broadly reflect a combination of climatic factors and plant distribution patterns, and had a banded, near-planar shape (Appendix A1) (Lee and Yim, 1978). Our SOM results, revealing four statistically significant spatial clusters (Zones I–IV) representing distinct territories (detailed below), were consistent with unification (or partial inclusion) or division of the historical zones. The historical central and southern zones partially reflected the distribution patterns of some tree ecotypes, including evergreen broad-leaved trees (e.g., *Citrus* and *Ilex*), but there was limited consideration of important biogeographic factors in small regions, such as complex mountain groups and the effects of altitude. Rather than forming a broad banded pattern, recent studies have revealed spatial patterns (at global and regional scales) that include large numbers of small, heterogeneous patches within larger biotic assemblage zones, arising from changes in the climatic and physical environment, such as quaternary glacial-interglacial oscillations, and the roles of factors such as topography (Kreft and Jetz, 2010; Lenormand et al., 2019; Médail and Diadema, 2009; Silva and Souza, 2018).

Zone I is focused around high-altitude mountains in the central Korean Peninsula (maximum altitude: 1708 m, Mt. Seoraksan) and included some high-altitude regions in the south. Among the historical zones, this partially corresponds to the central and southern zones. Zone I reflects floristic elements that moved south

from the previous periglacial environment and remained in the high-altitude mountain ranges in the center of the Korean Peninsula after the periglacial environment receded. In East Asia, this region represents the southern limit of the ranges of the dwarf Siberian pine (*Pinus pumila*), Korean arborvitae (*Thuja koraiensis*), and Khingan fir (*Abies nephrolepis*) (Kong et al., 2019; Korea National Arboretum, 2015). It is also an area with active cool-climate highland agriculture, and timber production (*P. koraiensis* and *Larix kaempferi*).

Zone II is focused on mountainous land adjacent to Zone I, mostly consisting of inland, high-altitude mountains further south in the Korean Peninsula (1915 m, Mt. Jirisan and 1614 m, Mt. Deogyusan). Among the historical zones, this partially corresponds to the same central and southern floristic zones as Zone I. This region is the northern limit of the range of the Korean fir (*A. koreana*), and the southern limit in East Asia of the range of dark-bark spruce (*Picea jezoensis*) (Korea National Arboretum, 2015). Through the same historical geographical processes as Zone I, this zone shows the remnants of the influence of the periglacial environment in the southern Korean Peninsula. In addition, trees that are mostly distributed in warmer or maritime climates (e.g., *Stewartia koreana* and *Lindera sericea*) can be found growing in the medium and high altitude areas of Zone II (Kim et al., 2014). There is almost no highland agriculture in this zone, but some timber production occurs (mostly *Larix kaempferi*), and the zone is adjacent to numerous large cities.

Zone III accounts for most of the southern Korean Peninsula. Although this zone contains tall mountains and mountain ranges, it is also subject to complex effects from agricultural activity and cities with a relatively intensive level of land use. For example, the inclusion of areas in Zone III with higher latitudes than Zone I is likely because of the impact of long-term military activity in these areas. Notably, in Zone III, the plants constituting the other floristic zones (Zones I, II, and IV) remain in isolated islands. These heterogeneous patches inside Zone III are a vestige of shifts in plant diversity patterns driven by historical geographical changes, and thus are important in the heterogeneous formation of the regional floristic composition (Laliberté et al., 2014; Zobel, 1997). Recently, the function of the Korean Peninsula as a shelter for biodiversity has been emphasized, because of its topography that includes a number of core mountains (Chung et al., 2018). In-depth studies need to be conducted on the biodiversity conservation functions (e.g., provision of shelter) of these small, heterogeneous patches and on their long-term changes.

Zone IV unifies the historical floristic zones of the southern coast, Jejudo, and Ulleungdo, which have previously been more finely divided (Appendix A1), and contains a high relative abundance of Orchidaceae, Asteraceae, and Rosaceae as well as specific taxa that only appear in this zone. Although Zone IV includes some inland areas, it mostly consists of regions along the coast of the Korean Peninsula and islands that are important for biodiversity, such as the oceanic island Ulleungdo (Chang and Gil, 2014; Choi et al., 2019; Holman et al., 2017) and the volcanic island Jejudo, which contains Hallasan Mountain (1950 m). This zone ranges from the temperate zone of evergreen broad-leaf trees (e.g., *Castanopsis sieboldii*) to the cool zone of polar trees (e.g., *Diapensia lapponica* var. *obovate*). Importantly, among the four identified floristic zones, Zone IV shows relatively high diversity of specific taxa, and is a core part of the range of Orchidaceae, making it an important zone from a conservation and evolutionary perspective. Because this zone includes the volcanic islands of Ulleungdo and Jejudo, the physical conditions differ considerably from the other zones, including in parent materials (volcanic ash) and terrain (lava terraces). The north of Jejudo has acted as an agricultural and administrative center for longer than the south, which could explain the differences in the flora.

Spatial clustering and separation of floristic assemblages

In addition to plant distribution, regionalization of floristic assemblages is important for understanding the formation processes on the Korean Peninsula, including floral changes. The regional species pool is the result of numerous settlement processes caused by climatic changes, which is restricted by ecological filtration, and ultimately forms the local biotic assemblage (Ricklefs, 1987). The Korean Peninsula is composed of a network of mountain ranges along a latitudinal gradient and has witnessed interactions between the Manchurian flora region at higher latitudes, the North Chinese flora region at lower latitudes, and the Japan-Korean flora

8

region (Takhtajan, 1986). This is the background for the current plant diversity and species composition on the peninsula (Appendix A2). Thus, through repeated historical geographical processes, such as periglacial climates, the Korean Peninsula has acted as a geographical and biological corridor, with a mixture of high- and low-latitude plants, which has produced the present-day spatial distribution of biodiversity (Chang et al., 2016; Chung et al., 2017a; Chung et al., 2017b; Kim et al., 2014; Kim et al., 2005; Kong et al., 2019).

Unlike the historical banded or homogeneous planar patterns, our revised floristic zones in the southern part of the Korean Peninsula reveal a new pattern, with Zones I and II forming central regions surrounded by a background of Zone III, with small patches of Zones I and II present within Zone III. At regional scales (e.g., the Korean Peninsula), complex physical and topographic factors can affect biotic assemblages (Lasmar et al., 2020; Tsiftsis et al., 2008; Xu et al., 2016). Across a broader area, there are also effects of climate, such as latitude (Sanders et al., 2007). The sequential settlement, expansion, contraction, and maintenance of species with historical geographic environmental oscillations are major processes involving interactions between the biota and topographic locations in a given region, which have been reported in other ecological regions (Lenormand et al., 2019; Silva and Souza, 2018). The Mt. Mudeungsan (1187 m) and Mt. Gyeryongsan (846 m) national parks in the center of the southern Korean Peninsula, the high-altitude Mt. Palgonsan (1192 m) region in the southeast Korean Peninsula, and the Youngnam Mountains (an aggregation of nine mountains over 1000 m) are important locations containing local-scale plant communities (e.g., *A. koreana* , *Primula farinose* , and *Carex tenuiformis* in the Youngnam mountains), which are distant from the heterogeneous flora and the central distributions of Zone III (Kim et al., 2015; Korea National Arboretum, 2014). The mountains in these regions are an important factor in the formation of heterogeneous, small-scale biological interactions (Gentili et al., 2015; Thomson, 2005) and the process of species differentiation.

Excluding the oceanic island Ulleungdo, the islands of the Korean Peninsula were last connected to the mainland before the Early Holocene, around 7,000 years ago (Kim et al., 2016). The islands have spent a long time, in terms of evolutionary biology, separated from the East Asian mainland. Among the zones defined in this study, the coastal regions and islands that constitute Zone IV (the warmest zone but with the least annual precipitation) showed especially high abundance of certain plant species (but low overall mean abundance) and specific plants (e.g., Orchidaceae and Asteraceae) were diverse. As discussed above, these results were likely because of the inclusion of the volcanic island Jejudo (1950 m above sea level) with its uniquely warm and humid maritime climate, and the oceanic island Ulleungdo (986.7 m above sea level). The Orchidaceae (71.4%)—e.g., *Habenaria chejuensis* —and Asteraceae (24.3%)—e.g., *Artemisia hallaisanensis* —that are only present in Zone IV include species that only grow on Jejudo or Ulleungdo. Orchidaceae are most abundant in warm and humid regions and show a negative correlation with the latitudinal gradient (Cribb et al., 2003; Myers et al., 2000). The family can be an index of biodiversity, since they show specialized habitat preference (Cho et al., 2019) and form associations between multiple species, including pollinators and mycorrhizal fungi (Pemberton, 2010). Spatial separation (as seen on isolated islands) is a core mechanism of species differentiation, but concomitant ecologically important environmental variables (e.g., climatic and physical conditions) also operate in combination with spatial separation to create selection pressures (Anacker and Strauss, 2014). Despite its narrow width, Zone IV is a key region in terms of ecologically important environmental characteristics, taxon diversity, and species differentiation, and thus, among the floristic zones on the Korean Peninsula, it occupies an important position from a conservation and evolutionary perspective.

Away from the coast, the mainland regions in Zone IV include Mt. Bukhansan, which is a national park close to the western coast, and Mt. Mudeungsan, which is in the southwest mainland. These regions are large mountains within Zone III and are thought to be vestiges left after the entry of coastal and island floristic features from a past environment. The slopes of Mt. Mudeungsan have many warm wind holes (Park, 2017) and contain numerous plants that can otherwise be found mostly along the coasts and in islands (e.g., *Cyrtosia septentrionalis* (Rchb. f.) Garay and Polypodiales) (Hong et al., 2013). Some islands to the south and west of the Korean Peninsula were categorized into Zone III., A landscape that includes grazing and crop farming predominates the islands in the south of the Korean Peninsula, and thus, there has been an active introduction of plant species from the mainland, changing the flora considerably over time (Kim et al., 2017). These ecological landscape characteristics could form the background for the high floristic connectivity

9

of some islands within Zone III, which mostly has flora in inland developed regions and warm temperate climates.

There are research limits to explaining in detail the processes and significance of spatial clustering and separation of floristic assemblages using regional patterns alone. Detailed descriptions of the plant reassembly process after the last glacial maximum and epoch-scale studies, such as pollen analysis, are needed (Yi, 2011). In addition, in the derived floristic zones, a convergent approach to phylogenetic history and diversity would be particularly useful for analyzing historical incidents in the formation of current biodiversity patterns, and for ascertaining the historical and evolutionary relationships between zones. Nevertheless, deducing the spatial arrangements of floristic assemblages is of great importance for understanding the ecologically important environmental factors involved in forming biogeographic regions. This study provides essential background knowledge to develop precise strategies for conservation based on micro- (Fenu et al., 2010) and nano-hotspots (Grant and Samways, 2011) at the local level. Moreover, there is great potential to quantitatively calculate the rarity, endemicity, and commonness of plant species, and to improve the priority of conservation and research (Casazza et al., 2005).

References

Ahn, H., Shin, J. Y., Jeong, C., & Heo, J. H. (2018). Assessing applicability of self organizing map for regional rainfall frequency analysis in South Korea. Journal of Korea Water Resources Association,**51** (5): 383–393.

Anacker, B. L., & Strauss, S. Y. (2014). The geography and ecology of plant speciation: range overlap and niche divergence in sister species. Proceedings of the Royal Society B: Biological Sciences,**281** (1778): 20132980.

Brum, F. T., Graham, C. H., Costa, G. C., Hedges, S. B., Penone, C., Radeloff, V. C., Rondinini, C., Loyola, R., & Davidson, A. D. (2017). Global priorities for conservation across multiple dimensions of mammalian diversity. Proceedings of the National Academy of Sciences,**114** (29): 7641–7646.

Casazza, G., Barberis, G., & Minuto, L. (2005). Ecological characteristics and rarity of endemic plants of the Italian Maritime Alps. Biological Conservation, **123** (3): 361–371.

Céréghino, R., Santoul, F., Compin, A., & Mastrorillo, S. (2005). Using self-organizing maps to investigate spatial patterns of non-native species. Biological Conservation, **125** (4): 459–465.

Chang, C. S., & Gil, H. Y. (2014). *Sorbus ulleungensis* , a new endemic species on Ulleung Island, Korea. Harvard Papers in Botany,**19** (2): 247–255.

Chang, C. S., Kim, H., Son, S., & Kim, Y. S. (2016). The red list of selected vascular plants in Korea. Korea National Arboretum and Korean Plant Specialist Group, Pocheon.

Chen, H., & Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics, **12** (1): 35.

Cho, Y. C., Kim, H. G., Koo, B. Y., & Shin, J. K. (2019). Dynamics and viability analysis of transplanted and natural lady's slipper (*Cypripedium japonicum*) populations under habitat management in South Korea. Restoration Ecology, **27** (1): 23–30.

Cho, Y. C., Kim, N. S., & Koo, B. Y. (2018). Changed land management policy and the emergence of a novel forest ecosystem in South Korea: landscape dynamics in Pohang over 90 years. Ecological Research,**33** (2): 351–361.

Choi, H. J., Yang, S., Yang, J. C., & Friesen, N. (2019). *Allium ulleungense* (Amaryllidaceae), a new species endemic to Ulleungdo Island, Korea. Korean Journal of Plant Taxonomy, **49** (4): 294–299.

Chon, T. S. (2011). Self-organizing maps applied to ecological sciences. Ecological Informatics, **6** (1): 50–61.

Chung, G. Y., Chang, K. S., Chung, J. M., Choi, H. J., Paik, W. K., & Hyun, J. O. (2017a). A checklist of endemic plants on the Korean Peninsula. Korean Journal of Plant Taxonomy, **47** (3): 264–288.

Chung, M. Y., López-Pujol, J., & Chung, M. G. (2017b). The role of the Baekdudaegan (Korean Peninsula) as a major glacial refugium for plant species: A priority for conservation. Biological Conservation,**206** : 236–248.

Chung, M. Y., Son, S., Suh, G. U., Herrando-Moraira, S., Lee, C. H., López-Pujol, J., & Chung, M. G. (2018). The Korean Baekdudaegan Mountains: A glacial refugium and a biodiversity hotspot that needs to be conserved. Frontiers in Genetic Science, **9** : 489.

Cottrell, M., Olteanu, M., Rossi, F., & Villa-Vialaneix, N. N. (2018). Self-organizing maps, theory and applications. Revista de Investigacion Operacional, **39** (1): 1–22.

Cribb, P. J., Kell, S. P., Dixon, K. W., & Barrett, R. L. (2003). Orchid conservation: a global perspective. In K. W. Dixon, S. P. Kell, R. L. Barrett, & P. J. Cribb (Eds.), *Orchid Conservation* (pp. 1–25). Sabah Natural History Publications, Kota Kinabalu.

Fenu, G., Mattana, E., Congiu, A., & Bacchetta, G. (2010). The endemic vascular flora of Supramontes (Sardinia), a priority plant conservation area. Candollea, **65** (2): 347–358.

Gentili, R., Bacchetta, G., Fenu, G., Cogoni, D., Abeli, T., Rossi, G., Salvatore, M. C., Baroni, C., & Citterio, S. (2015). From cold to warm-stage refugia for boreo-alpine plants in southern European and Mediterranean mountains: the last chance to survive or an opportunity for speciation? Biodiversity, **16** (4): 247–261.

Giraudel, J. L., & Lek, S. (2001). A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. Ecological Modelling, **146** (1–3): 329–339.

Gonzalez-Orozco, C. E., Ebach, M. C., Laffan, S., Thornhill, A. H., Knerr, N. J., Schmidt-Lebuhn, A. N., Cargill, C. C., Clements, M., Nagalingum, N. S., Mishler, B. D., & Miller, J. T. (2014). Quantifying phytogeographical regions of Australia using geospatial turnover in species composition. PloS One, **9** (3): e92558.

Graham, C. H., & Hijmans, R. J. (2006). A comparison of methods for mapping species ranges and species richness. Global Ecology and Biogeography, **15** (6): 578–587.

Grant, P. B., & Samways, M. J. (2011). Micro-hotspot determination and buffer zone value for Odonata in a globally significant biosphere reserve. Biological Conservation, **144** (2): 772–781.

Holman, G., Del Tredici, P., Havill, N., Lee, N. S., Cronn, R., Cushman, K., Mathews, S., Raubeson, L., & Campbell, C. S. (2017). A new species and Introgression in Eastern Asian Hemlocks (Pinaceae: Tsuga). Systematic Botany, **42** (4): 733–746.

Hong, H. H., Jang, J. W., Sun, E. M., Kim, B. A., Kim, S. J., Seo, S. R., & Im, H. T. (2013). Floristic study of Mt. Mudeung. Environmental Biology Research, **31** (2): 121–153.

Ibanez-Erquiaga, B., Pacheco, A. S., Rivadeneira, M. M., & Tejada, C. L. (2018). Biogeographical zonation of rocky intertidal communities along the coast of Peru (3.5-13.5 degrees S Southeast Pacific). PLoS One, **13** (11): e0208244.

Kim, H. H., Kim, D. B., Won, H. K., Kim, C. S., & Kong, W. S. (2016). Island-biogeographical characteristics of insular flora in Southern Sea of Jeollanamdo, Korea. Journal of Climate Change Research,**7** (2): 143–155.

Kim, H. H., Kim, D. B., Jeon, C. H., Kim, C. S., & Kong, W. S. (2017). Island-biogeographical characteristics of naturalized plant in Jeollanamdo islands. Journal of Environmental Impact Assessment,**26** (4): 273–290.

Kim, J. S., Chung, J. M., Kim, S. Y., Kim, J. H., & Lee, B. Y. (2014). Phytogeographic study on the Holocene hypsithermal relict plant populations in the Korean Peninusula. Korean Journal of Plant Taxonomy,**44** (3): 208–221.

Kim, N. S., Han, D., Cha, J. Y., Park, Y. S., Cho, H. J., Kwon, H. J., Cho, Y. C., Oh, S. H., & Lee, C. S. (2015). A detection of novel habitats of Abies Koreana by using species distribution models (SDMs) and its application for plant conservation. Journal of the Korea Society of Environmental Restoration Technology, **18** (6): 135–149.

Kim, Z. S., Hwang, J. W., Lee, S. W., Yang, C., & Gorovoy, P. G. (2005). Genetic variation of Korean pine (*Pinus koraiensis* Sieb. Et Zucc) at allozyme and RAPD markers in Korea, China and Russia. Silvae Genetica, **54** (1–6): 235–246.

Kira, T. (1945). A new classification of climate in eastern Asia as the basis for agricultural geography. Kyoto, Japan: Horticultural Institute, Kyoto University.

Kohonen, T. (1995). Self-Organizing Maps. Springer, Berlin, pp 502.

Kong, W. S., Kim, H., Kim, D., Lee, C. H., & Shin, H. T. (2019). Natural History and Climate Change of Dwarf Relic Shrubs. Advan Publishing Inc., Seoul.

Korea National Arboretum (2014). Conservation of Korean fir (*Abies koreana*) in a changing environment. Forest of Korea, I. Sumeungil, Seoul.

Korea National Arboretum (2015). Conifers. Ecology of woody plants in South Korea, I. Sumeungil, Seoul.

Korea National Arboretum (2016). Distribution maps of vascular plants in Korea. Korea National Arboretum, Pocheon.

Korean Meteorological Administration, K. (2020). Weather information. https://data.kma.go.kr, accessed January 1, 2020.

Kreft, H., & Jetz, W. (2010). A framework for delineating biogeographical regions based on species distributions. Journal of Biogeography, **37** (11): 2029–2053.

Laliberté, E., Zemunik, G., & Turner, B. L. (2014). Environmental filtering explains variation in plant diversity along resource gradients. Science, **345** (6204): 1602–1605.

Lasmar, C. J., Ribas, C. R., Louzada, J., Queiroz, A. C., Feitosa, R. M., Imata, M. M., Alves, G. P., Nascimento, G. B., Neves, F. S., & Domingos, D. Q. (2020). Disentangling elevational and vegetational effects on ant diversity patterns. Acta Oecologica, **102** : 103489.

Lee, W. C., & Yim, Y. J. (1978). Studies on the distribution of vascular plants in the Korean peninsula. Korean Journal of Plant Taxonomy, **8** (Appendix): 1–33.

Lenormand, M., Papuga, G., Argagnon, O., Soubeyrand, M., De Barros, G., Alleaume, S., & Luque, S. (2019). Biogeographical network analysis of plant species distribution in the Mediterranean region. Ecology and Evolution, **9** (1): 237–250.

Linder, H. P., de Klerk, H. M., Born, J., Burgess, N. D., Fjeldså, J., & Rahbek, C. (2012). The partitioning of Africa: statistically defined biogeographical regions in sub-Saharan Africa. Journal of Biogeography, **39** (7): 1189–1205.

Liu, K., Kargupta, H., & Ryan, J. (2006). Random projection-based multiplicative data perturbation for privacy preserving distributed data mining. IEEE Transactions on Knowledge and Data Engineering, **18** (1): 92–106.

McLaughlin, S. P. (1989). Natural floristic areas of the western United States. Journal of Biogeography, **16** : 239–248.

Médail, F., & Diadema, K. (2009). Glacial refugia influence plant diversity patterns in the Mediterranean Basin. Journal of Biogeography, **36** (7): 1333–1345.

Ministry of Land Infrastructure and Transport (2016). The National Atlas of Korea II. National Geography Information Institute, Suwon.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. Nature, **403** (6772): 853–858.

Nakai, T. (1919). Report on the vegetation of the Island Ooryongto or Dagelet Island, Corea. The Government of Chosen, Seoul, Korea.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., & Loucks, C. J. (2001). Terrestrial ecoregions of the world: A new map of life on Earth. BioScience, **51** (11): 933–938.

Park, C. W. (2017). A study on the characteristics of warm wind hole zone of talus slope in Mt. Mudeung National Park. Journal of the Association of Korean Geographers, **6** (3): 381–393.

Pemberton, R. W. (2010). Biotic resource needs of specialist orchid pollinators. The Botanical Review, **76** (2): 275–292.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ricklefs, R. E. (1987). Community diversity: relative roles of local and regional processes. Science, **235** (4785): 167–171.

Ricklefs, R. E. (2004). A comprehensive framework for global patterns in biodiversity. Ecology Letters, **7** (1): 1–15.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, **20** : 53–65.

Rural Development Administration, R. (2010). Korean Soil Information System.

Sanders, N. J., Lessard, J. P., Fitzpatrick, M. C., & Dunn, R. R. (2007). Temperature, but not productivity or geometry, predicts elevational diversity gradients in ants across spatial grains. Global Ecology and Biogeography, **16** (5): 640–649.

Silva, A. C., & Souza, A. F. (2018). Aridity drives plant biogeographical sub regions in the Caatinga, the largest tropical dry forest and woodland block in South America. PLoS One, **13** (4): e0196130.

Snedden, G. A. (2019). Patterning emergent marsh vegetation assemblages in coastal Louisiana, USA, with unsupervised artificial neural networks. Applied Vegetation Science, **22** (2): 213–229.

Stoddart, D. R. (1992). Biogeography of the tropical Pacific. Pacific Science, **46** (2): 276–293.

Sun, Y., Yan, X., & Xie, D. (2008). A new method of vegetation–climate classification in China. International Journal of Climatology,**28** (9): 1163–1173.

Takhtajan, A. (1986). Floristic Regions of the World. California Press, Berkeley, pp. 522.

Thomson, J. D. (2005). Ecological differentiation of Mediterranean endemic plants. Taxon, **54** (4): 967–976.

Tsiftsis, S., Tsiripidis, I., Karagiannakidou, V., & Alifragis, D. (2008). Niche analysis and conservation of the orchids of east Macedonia (NE Greece). Acta Oecologica, **33** (1): 27–35.

Valladares, F., Bastias, C. C., Godoy, O., Granda, E., & Escudero, A. (2015). Species coexistence in a changing world. Frontiers in Plant Science, **6** : 866.

Vilhena, D. A., & Antonelli, A. (2015). A network approach for identifying and delimiting biogeographical regions. Nature Communications, **6** (1): 1–9.

Wehrens, R., & Kruisselbrink, J. (2018). Flexible self-organizing maps in kohonen 3.0. Journal of Statistical Software, 87(7): 1–18.

White, F. (1983). The Vegetation of Africa. UNESCO, Paris, pp. 356.

Whittaker, R. H. (1962). Classification of natural communities. The Botanical Review, **28** (1): 1–239.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York, pp. 224.

Xu, X., Zhang, H., Tian, W., Zeng, X., & Huang, H. (2016). Altitudinal patterns of plant diversity on the Jade Dragon Snow Mountain, southwestern China. Springerplus, **5** (1): 1566.

Yi, S. (2011). Holocene vegetation responses to East Asian monsoonal changes in South Korea. In J. Blanco & H. Kheradmand (Eds.), Climate Change - Geophysical Foundations and Ecological Effects (pp. 154–178). InTech, Rijeka.

Yim, Y. J., & Kira, T. (1975). Distribution of forest vegetation and climate in the Korean peninsula. I. Distribution of some indices of thermal climate. Japanese Journal of Ecology, **25** (2): 77–88.

Zar, J. H. (1984). Biostatistical analysis, Englewood Cliffs. NY: Prentice-Hall, 360.

Zobel, M. (1997). The relative of species pools in determining plant species richness: an alternative explanation of species coexistence? Trends in Ecology & Evolution, **12** (7): 266–269.

Figure legends

Fig. 1. Representativeness of study sites. (a) Distribution of study sites across the southern part of the Korean peninsula. (b) Distribution of study sites across the global vegetation biomes defined by the Whittaker classification.

Fig. 2. Conceptual diagram of the analysis steps and data flow in defining the floristic zonation (Kreft and Jetz, 2010).

Fig. 3. Results of the self-organizing map (SOM) analysis for 9 × 9 SOM map. (a) The neighbor distance plot or U-Matrix indicating the distance between each node and its neighbors. (b) The classification of the training samples according to the SOMs.

Fig. 4. Four floristic zones mapped to the southern Korean Peninsula, derived using plant distribution data. The straight and solid lines represent the named historical floristic regions. The lines dividing the historical floristic regions have previously been used to explain the observed distribution of plants on the Korean Peninsula (e.g., north, central, and southern Korean Peninsula).

Fig. 5. Comparison of geographic and climate factors in the derived floristic zones on the southern Korean Peninsula. (a) latitude; (b) longitude; (c) mean annual temperature; (d) annual precipitation; (e) warmth index; and (f) coldness index. Median values are marked inside the boxplots, errors bars indicates the 90th and 10th percentiles, and the points outside of the error bars indicate outliers. A Tukey post-hoc test revealed significance groups, represented by letters ($p < 0.05$).

Fig. 6. Analysis of composition ratios for physical factors in each floristic zone on the Korean Peninsula. (a) Parent material; (b) topography; (c) effective soil depth; and (d) soil texture.

Fig. 7. A Venn diagram showing the relationships between plant distributions in each of the four floristic zones. Numbers represent individual taxonomic groups.

Fig. 8. Comparison of the composition of specific plant families in each floristic zone. As shown in Fig. 7, there were 1,099 taxa that appeared in all 4 zones. The specific plant species in Zones I, II, III, and IV numbered 72, 25, 192, and 404 taxa, respectively.

Data Accessibility Statement

Competing Interests Statement

The authors declare that they have no competing interests.

Author Contributions

Song Hie Jung: Writing—original draft (lead); formal analysis (lead); methodology (equal). Yong-chan Cho: Conceptualization (lead); methodology (equal); writing—review and editing (lead).

Acknowledgements

Tables

Table 1. The geographical and biological statistics used to delineate the four floristic zones of the southern Korean Peninsula.

Floristic zones	Area (km^2) (%)	Species richness	Mean species richness (per grid) ($\pm$ S.E.)
Zone I (Cold floristic zone in central high elevation areas)	12,419 (12.3)	1,700	464.7 $\pm$ 9.9
Zone II (Cool floristic zone in central and southern high elevation areas)	11,619 (11.5)	1,668	331.7 $\pm$ 11.5
Zone III (Warm floristic zone in central and southern hilly areas)	72,367 (72.0)	2,379	79.3 $\pm$ 3.4
Zone IV (Maritime floristic zone near the coast and islands)	4,265 (4.2)	2,200	298.6 $\pm$ 14.0
Total	100,378 (100.0)	2,954	162.6 $\pm$ 5.8

Table 2. Analysis of the composition ratios (%) of physical environmental factors in the research area and derived floristic assemblage zones, including topography, soil parent materials, soil depth, and soil type.

Parameters	Zone I	Zone II	Zone III	Zone IV	Total land area
Topological classes					
Mountain	44.6	69.2	36.8	35.5	46.9

15

Parameters	Zone I	Zone II	Zone III	Zone IV	Total land area
Hill	21.4	8.8	23.1	12.2	18.2
Pediment	14.7	7.6	15.8	7.0	12.8
Interrill area	7.6	9.0	7.4	11.3	8.1
Fan	2.5	2.1	2.4	2.7	2.3
Lava terrace	-	0.2	0.5	20.1	1.5
Others	9.0	3.1	14.0	11.2	10.1
Sum	100.0	100.0	100.0	100.0	100.0
Parent material types					
Acidic rocks	50.1	46.3	48.2	56.9	48.5
Metamorphic rocks	26.2	38.2	22.4	5.0	26.3
Sedimentary rocks	13.5	11.5	13.8	0.7	12.3
Quaternary deposits	9.8	3.7	14.6	10.2	10.6
Volcanic ash	-	-	0.4	23.7	1.7
Others	0.3	0.3	0.6	3.4	0.6
Sum	100.0	100.0	100.0	100.0	100.0
Soil depth classes (cm)					
<20	19.0	20.9	17.7	26.8	19.3
20–50	21.6	27.4	20.2	25.4	22.7
50–100	40.5	42.9	38.4	24.9	39.1
>100	18.9	8.9	23.7	22.9	18.8
Sum	100.0	100.0	100.0	100.0	100.0
Soil texture types					
Sandy gravel	2.2	1.4	2.5	4.4	2.3
Silt sandy loam	52.6	69.4	44.5	27.8	51.6
Clay loam	41.5	27.1	47.9	63.1	42.1
Clay	3.7	2.2	5.2	4.7	4.1
Sum	100.0	100.0	100.0	100.0	100.0

16
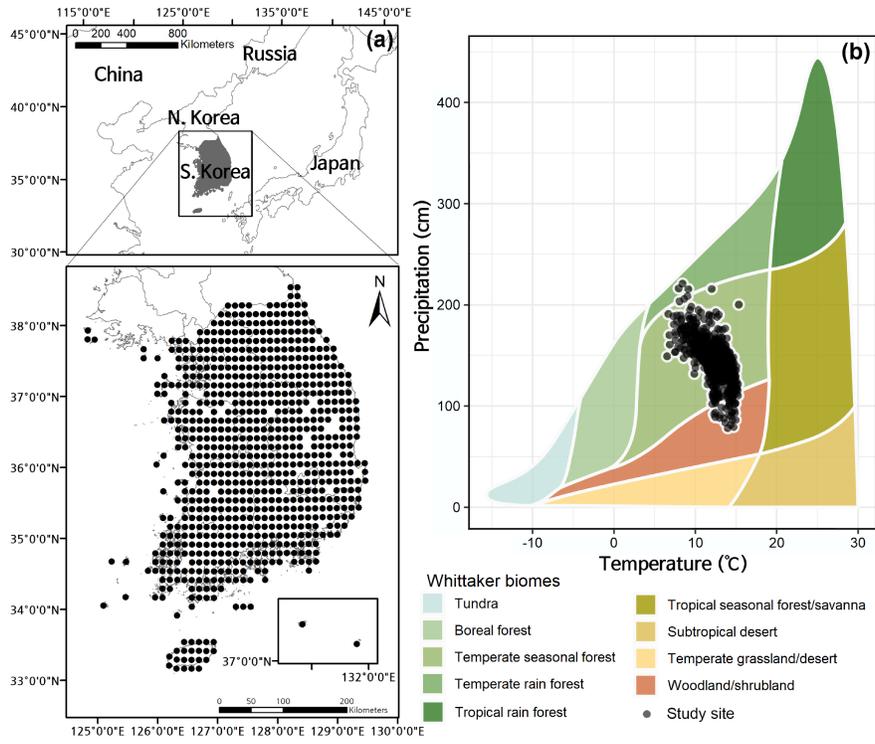
[1] Constructing species distribution database

[2] Determinating grid size and Intergrating point data into grid units

Presence/absence data
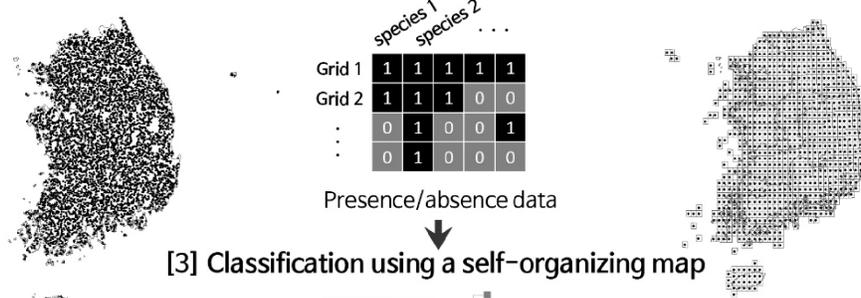
[3] Classification using a self-organizing map
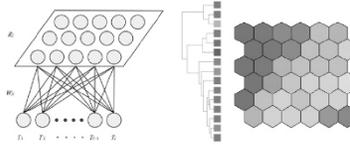
[4] Cluster analysis

[5] Mapping and categorical spatial interpolation

Map of floristic zones

Zone II

Zone I

Zone III

Zone IV

25

72

45

31

12

192

254

13

404

1099

135

19

84

182

388

Apiaceae (4)
Asteraceae (5)
Rosaceae (4)
Caryophyllaceae (3)
Fabaceae (5)
Orchidaceae (3)
Woodsiaceae (2)
Ranunculaceae (6)
Papaveraceae (2)
Crassulaceae (2)
Grossulariaceae (2)
Gentianaceae (2)
Rubiaceae (2)
Cyperaceae (7)
Liliaceae (2)

Zone I
(Cold floristic zone)

Asteraceae (4)
Scrophulariaceae (4)
Polygonaceae (2)
Liliaceae (2)
Rosaceae (4)
under one species in genus (9)

Zone II
(Cool floristic zone)

Rosaceae (47)
Fabaceae (55)
Liliaceae (47)
Cyperaceae (68)
Polygonaceae (32)
Apiaceae (32)
Lamiaceae (30)
under one species in genus (20)
Ranunculaceae (29)
Poaceae (98)
Brassicaceae (26)
Violaceae (19)
Dryopteridaceae (18)
Scrophulariaceae (17)
Orchidaceae (17)
Caryophyllaceae (16)
Caprifoliaceae (15)
Salicaceae (12)
Celastraceae (12)
Athyriaceae (11)
Asteraceae (113)
Rubiaceae (11)
Convolvulaceae (11)

All zones

under ten species in genus (363)

Rosaceae (8)
Brassicaceae (10)
Polygonaceae (7)
Asteraceae (10)
Ranunculaceae (7)
Poaceae (18)
Scrophulariaceae (7)
Liliaceae (5)
Orchidaceae (5)
Fabaceae (5)
Cyperaceae (19)
Caryophyllaceae (4)
Chenopodiaceae (4)
Papaveraceae (4)
Lamiaceae (4)
Solanaceae (4)
Amaranthaceae (3)
Crassulaceae (3)
Primulaceae (3)
Oleaceae (3)
Verbenaceae (3)
Salviniaceae (3)
under two species in genus (53)

Zone III
(Warm floristic zone)

Cyperaceae (22)
Rosaceae (21)
Dryopteridaceae (14)
Athyriaceae (12)
Liliaceae (12)
Asteraceae (27)
Fabaceae (12)
Orchidaceae (27)
Polypodiaceae (11)
Lamiaceae (10)
Scrophulariaceae (10)
Ranunculaceae (9)
Aspleniaceae (8)
Rubiaceae (8)
Oleaceae (7)
under five species in genus (131)
Polygonaceae (7)
Urticaceae (6)
Violaceae (6)
Caprifoliaceae (6)
Campanulaceae (6)
Brassicaceae (6)

Zone IV
(Maritime floristic zone)

21