Rapid Botanic Survey, Bioquality and improving botanical inventory in the tropics by integrating across spatial scales

W. D. Hawthorne[1] & C. A. M. Marshall[2]

[1]Department of Plant Sciences, South Parks road, Oxford, OX1 3RB

will.hawthorne@btopenworld.com (corresponding author)

[2]Department of Plant Sciences, University of Cambridge, Cambridge, CB2 3EA, UK

Running title: Rapid Botanic Survey and Bioquality

ABSTRACT. We review the Rapid Botanic Survey method (RBS), in the context of botanical recording to date. The concept of bioquality, a biodiversity value respecting global rarity, is summarised. Bioquality assessment involves the Star system for categorising species by global rarity; and a Genetic Heat Index (GHI) which aggregates Stars into community scores. All vascular plant species in tropical Africa have Stars, and >3.1 million botanical records have been databased across the continent (Marshall et al. 2016). Presented here are updated bioquality scores from continental tropical Africa, and especially coastal East Africa, and calculated for sample units of various shapes and sizes: East African Flora regions, one degree squares, forest reserves, to fine scale (sample-level) hotspots along the East African coast. GHI is globally standardised and has been calculated for survey data outside Africa, though seldom in tropical Asia. RBS data can also be used to distinguish vegetation types and can include ethnobotanical data. It is recommended as a way to standardise biodiversity or environmental Impact assessment nationally and globally, and for integrating such survey results in databases that will be increasingly useful as the tension between conservation and deforestation increases, and the climate changes.

Keywords. Africa, bioquality, endemism, hotspot, plants, scale, Rapid Botanic Survey

Introduction

David Mabberley's influence

The first author's undergraduate introduction to tropical botany, and then D.Phil thesis in the early 1980s, was supervised by David Mabberley (also by Frank White, curator of Oxford herbaria, and by Barrie Juniper). WH's

thesis, on the coastal forests of East Africa, involved several lines of research that continue today, including: tropical vegetation survey (mainly Africa); botanical inventory; and forest ecology as a basis for sustained use management and conservation of plant diversity. Underlying these has been the development of the Rapid Botanic Survey and bioquality approach (RBS), described below. RBS requires identification of plants of all ages, sizes and plant families, so field guide production and identification of tropical plants from sterile material has become a research theme in itself. The original thesis started as an ethnobotanic expedition to (sacred) kaya forests of the Kenyan coast, and RBS still includes ethnobotanic survey where appropriate. The kaya forests are revisited in this paper. Digital and databased botanical data is now an essential component of RBS, and although the D.Phil. was in the early days of botanical computing, a start was made down this avenue as well. The second author completed her D.Phil in Oxford's Plant Science Department under the supervision of the first, and now continues her botanic research in Cambridge's Plant Science Department. CM's research has continued in the tradition, whilst introducing more data-intensive, statistical and modelling approaches.

Botanical Sampling

A fundamental particle in the world of botanical data is a well-curated and well-localised botanical record, preferably linked to a herbarium specimen, fixing plants in defined places at a given time. Based on these primary data, secondary botanical records are generated, in Floras, monographs, checklists and others. Data clouds composed of digitised botanical records reveal patterns and processes from global to metre scale: understanding of these patterns and processes is crucial for management of plant resources.

Primary botanical records can be divided into taxonomic or ecological types. Within the classical, more taxonomic tradition, plant explorers select fertile specimens of plants they come across, often serendipitously, particularly if they are unknown species or new records for the region. Collectors may make ecological notes about the plant community, and indeed the short essays and 'growing with' species lists on the labels of the best-documented herbarium specimens were one of the inspirations for RBS approach, described below. Whilst these endeavours have provided the foundation for botany as a whole, our knowledge of the distribution of plant life after centuries of botanical recording effort in the tropics is strongly biased, with botanical records concentrated near to roads, rivers or the coast and with difficult-to-press, less conspicuous or accessible types of plant generally under-represented (Stropp et al., 2016). Curators prevent commoner plants from becoming proportionately common in herbarium cupboards, and older (pre-GPS or even pre-map) records are often poorly localised or relate to vegetation that has been long destroyed. Individual collectors collect only species of their preferred families. Many taxonomical collectors put most of their efforts into seeking and collecting new or rare species and consider records of other species relatively worthless, particularly if sterile: the botanical world loses some interest after the fanfare of a new species' arrival.

In the more ecological tradition meanwhile, ecologists, foresters, biodiversity- and carbon-assessors record all plants or species in or along a carefully or randomly selected square, circle or measured line, at a particular point in time, sometimes returning regularly to the same area to collect updates (e.g. Hawthorne et al., 2012). They may collect voucher specimens to allow identification later, but the main aim is to describe the population or plant community. Typically, the same species is recorded many times in the same sample plot. Although taxonomic precision is not always given as much attention as it deserves, species from many families are covered, though there is usually a limit to the types of plant surveyed (trees only, for instance). Surveys with

plots that cover whole countries systematically — national forest inventories — usually include only tree species, especially in tropical regions. Systematic, region-wide surveys in the tropics of all types of vascular plant species in a given timeframe are too few, although examples include Hall and Swaine (1981), Hawthorne & Abu Juam (1995), Howard et al. (2000), Baksh-Comeau et al. (2016), and usually focus on protected areas, with secondary vegetation and small patches of primary vegetation under-sampled. The shortage is partly due to the perceived high costs of scaling up the efforts required for single measured plots. Thorough, taxonomically careful checklists of small areas, perhaps compiled from a concerted effort to summarise herbarium data and revisit and collect specimens from a given forest or park, are very useful at filling in the data gaps, but are similarly too few and may not be very highly resolved spatially (e.g. Meerts, 2016; Lindsell et al., 2019).

All this leads to a very haphazard record of the plant life across the planet. A complete coverage of thorough checklists for small homogenous habitat units across the landscape would be ideal: the aim of RBS is to achieve this by combining the strengths of both taxonomic and ecological recording, in a bid to provide useful data on all species of (vascular) plant, the types of plant community, and the location of botanical hotspots in a landscape or country. The general purpose of RBS is to enable better decisions over the stewardship of dwindling plant biodiversity in an increasingly populated landscape and changing environment; also to document it for posterity. Without taxonomic precision we have inefficient environmental impact assessors, missing the rarer plants in the area about to be flooded, mined or built upon; and without completeness of local samples, the meaning of regional patterns is less clear. In common with field survey methods (relevés) associated with Braun-Blanquet phytosociology, sampling depends on subjective, surveyor's choice of representative sample sites (Westhoff & van der Maarel, 1978; Wikum & Shanholtzer, 1978). The relevé method was designed for European types of vegetation, is not ideal for rainforest, and the relevés are measured, unlike typical RBS samples.

An RBS field sample aims to record all, or at least most species in a defined landscape unit, in up to half a day's collecting by a small team of 4-6 (see Hawthorne & Marshall, 2016 for a manual). Most plants are inevitably sterile, and the default tactic is to collect single specimens of at least foliage of all species. A minority, the very common and well known species, may be observed only, particularly if vouchered earlier in the same region. The protocol is mainly concerned with making as efficient as possible the drying, sorting and managing of the industrial quantities of dried plant material (up to 300 specimens a day per team, though only a few are ultimately acceded to formal herbarium collections). The identification of sterile, sometimes juvenile specimens is facilitated by sorting thousands of these specimens at a time, producing a chain of morphotypes within each species, some part of which can be closely matched to a fertile herbarium specimen.

Measuring plant hotspots and bioquality

Identifying areas of high biodiversity is an established way to prioritize areas for conservation (e.g. Stattersfield et al. 1988; Myers et al., 2000; Olson & Dinerstein, 2002). 'Biodiversity hotspots' were originally identified using the richness of species endemic to large, biogeographic realms which had been significantly degraded (Myers et al., 2000), in part because species distribution data were available only at this coarse resolution. Richness hotspots, with high numbers of species (per unit area), may highlight areas with greater topographic and climatic diversity; or merely indicate areas that are biogeographic and ecological crossroads, part of a large, extensive species pool, with widespread species freely migrating from adjacent areas (Barthlott et al., 2005). Very isolated oceanic or

habitat islands may have a low richness, yet high biodiversity significance. The calculation of reliable richness metrics requires complete datasets or reliance on uncertain extrapolations from species/sampling curves, and the herbarium collection dataset is far from statistically ideal for such extrapolation. Thus, richness alone is now generally considered to have limited value for highlighting priorities for conservation, or for identifying significant areas in the context of evolutionary history (Brooks et al., 2006).

Indices of endemicity — the number or proportion of rare or range-restricted species in a plant community or flora — are usually better indicators of areas deserving greater conservation effort. A proportional index such as percent endemism may be calculated from incomplete enumerations of species, although bias may arise where collectors have focused on either the commoner or rarer plants. The recognition of a species as an endemic depends on defining a focal area, for instance a country, mountain or other arbitrary area of any size (e.g. $<50,000$ km^2, IFC Performance Standard 6). In the range size rarity approach, species are scored in inverse proportion to the number of cells occupied in a grid drawn over a region, and grid cell scores are derived as the sum or mean of the weighted species scores within them (e.g. Williams, 1996; Kier & Barthlott, 2001). This goes some way to standardising 'endemism', but in its typical application, this approach considers rarity only within the boundaries of the study extent. The principle of focussing on global range of species is seldom disputed when seeking the global significance of particular plant communities or regions, but it is currently impossible to make a precise range size rarity index based on a global grid, because different regions are sampled so inconsistently.

These considerations led to the development of the Rapid Botanic Survey (RBS) and bioquality approach to the assessment of priority areas and impacts of disturbance on global biodiversity (Hawthorne & Marshall, 2016), initially for application within Ghana's Forest reserve system (Hawthorne & Abu Juam, 1995).

'Bioquality' is defined as the biodiversity value of a plant community that is due to the concentration of globally rare taxa in that community, where the value ascribed to a community is weighted by the degree of global rarity (geographical range size) of the component species (Hawthorne & Abu Juam, 1995; Hawthorne 1996). Bioquality is typically measured using the Genetic Heat Index (GHI). If unspecified, GHI scores have been calculated from the whole vascular plant community. Alternatively, sets of species may be specified, for example 'tree community bioquality' might be calculated as the GHI for tree species with individuals above a certain specified size class, within a sampled area. Like biodiversity, bioquality is a property of a set or community of species. Bioquality' can be considered alongside other parameters of habitat quality such as forest type, richness, diversity, rarity in a local context, community complementarity and ecosystem service value (Hawthorne, 1996).

Unlike measures of diversity (such as Fisher's alpha, Simpson's D or the Shannon-Weaver index) which characterise the number of species and their relative abundance, bioquality characterises the concentration of valued species in a sample or community. Vegetation on an old, urban rubbish tip can have a high diversity, with many different exotic and pioneer species, but if these species are all globalised weeds, the bioquality would be very low. Conversely, a patch of forest dominated by hundreds of trees of one locally endemic tree species in a mono-specific genus would have higher bioquality than a patch of similar forest with 100 widespread species. Every individual plant in the first case is part of a high priority, globally rare species, and in the second case there are no individuals of globally rare species. One expects, in general, to encounter globally rare plants (and genes) more frequently whilst walking randomly in high bioquality vegetation than in low bioquality vegetation.

There is no point in being over-precise in the measurement of bioquality, given our highly incomplete knowledge of global plant distribution. Yet it is still possible to rank plant communities objectively, based on a crude index of global rarity represented by their component plant species. The signal obtained by combining, by weighted average, the statistically noisy global rarity scores for tens or hundreds of species in a plant community varies subtly in a way that can be explored and modelled statistically. This works as surely as counting in a busy room the number of people classified into crude age classes using observable criteria (young, middle-aged and old); and then estimating from other sources the actual average age of people in the general population classified by the same criteria; and finally calculating the weighted average (N.people in each class x weight (=av. Age) of people of that class). Imprecise, but with a large enough sample size, this can be used to classify or rank crowds adequately in different venues into younger or older age profiles, on a continuous scale.

Bioquality is a general term, like temperature. A particular scale or index, GHI (Genetic Heat Index) is to bioquality what centigrade is to temperature. To calculate GHI for a sample of species, all species are put in a category — Black, Gold, Blue or Green Star — based on an assessment of their global range. Range size is measured as the number of degree squares occupied by the species, with the degree square grid origin at the equator-meridian intersect. In estimating a species' global range, all available sources are used, including Floras; a statement like 'Widespread in the tropics' would earn a species the most widespread category of Green Star, whereas a statement like 'endemic to Mt. Nuang' would earn a species the rarest category, Black Star. Increasingly, online databases of botanical records are used to help make these decisions, though with caution. It is easy to recognise many species as Green Star species (widespread species), even based on lists of countries occupied as in Kew's World Checklist of Selected Plant Families (Govaerts et al., 2019); and at the other extreme, many species can be immediately classified as Black Star, when they are endemic to a small area (typically present in < 4 degree squares). The intermediate Blue and Gold Star categories take more effort to discriminate. Stars have some use in their own right, when shortlisting rarer species for IUCN assessment for instance (Baksh-Comeau et al., 2016), but are mainly used for calculating GHI.

The weights for the Stars are, for tropical and subtropical species, in inverse proportion to the mean number of degree squares known for a subset of species within them, especially those that have been mapped in recent monographs. In order to align the definition of Stars around the world, there are average degree square ranges expected for each of the Stars (Table 1): in any project region, different rules are developed to classify species to attain the standard mean global ranges. Outside the tropics, 100 km grid cells are used instead of degrees. The system involves many other considerations, for example species known to be rare within their range can be upgraded, and infraspecific taxa can in some circumstances be upgraded relative to the species as a whole. But the basic principle is simple, and the process is very rapid compared to, for instance, IUCN Red listing (IUCN, 2019).

GHIs for plant communities, grid cells or regions are then calculated as weighted averages of the Stars' weights, for the species present within them. The actual range of plant species is too poorly known to make it worth relying on species-specific range sizes and weights. Nevertheless, the crude scores aggregated from many species yields a useful, subtle index that relates well to gradients in and around hotspots (Marshall et al., 2016; Baksh-Comeau et al., 2016).

There are two main features of GHI in favour of its use an index of bioquality, and of biodiversity value in a global context: GHIs can be compared globally and this can for instance allow conservation organisations to allocate conservation payments by reference to it (Hawthorne et al., 1998). Secondly, GHI can be calculated at various scales (like other proportional values, including pH and ordination scores). Hence, sampling of local communities need not be constrained to a particular shape or size of a plot facilitating, as in the RBS approach, pure samples for habitats that are otherwise difficult to sample, like cliffs, winding stream banks or scattered termite hill vegetation amalgamated in savanna. The species list need only be a fair summary for the stated landscape unit, habitat type or region (fair at least with respect to the balance of Stars). As a rule of thumb, the sample should contain more than 40 species, and it is advisable to compare between samples at the same scale: it does, however, allow investigation of the impact of scale of assessment on the perceived bioquality. GHI reveals interesting local hotspots across landscapes, important knowledge in an era of ever-increasing pressure for multiple fine scale uses.

Aims

We demonstrate for East Africa the potential value, and shortage, of data at an appropriate resolution for good management decisions, as well as the value of the RBS and bioquality framework for filling these gaps efficiently and assessing global biodiversity priorities at all scales. We suggest RBS would be a useful method anywhere in the world. We finish by highlighting the uses to which RBS datasets can be put beyond bioquality assessment.

Methods

A database describing tropical African vascular plant species distributions was compiled from available records within a standard taxonomic framework (see Marshall et al., 2016 for details). 3.1 million global species distribution records were assembled for tropical African vascular plants, from plot data, herbarium databases, checklists, Floras, monographs and the Global Biodiversity Information Facility (GBIF). We limited our GBIF search to records supported by herbarium specimens and those without reported geographic issues. Our tropical African species list was derived from the African Plants Database (2014), and includes 40,583 accepted species or intraspecific names which were checked for synonymy and comprehensiveness against other resources. Updates and additions have been added subsequently, an ongoing process. We summarised the global range for all species into Stars, using the method outlined in the introduction and described fully in Hawthorne & Marshall (2016). Stars are updated as taxonomy evolves, and as new records change apparent range sizes.

The database was managed initially with the Brahms 7 software (Department of Plant Sciences, 2013), and has subsequently been managed in customised Visual Foxpro database designed by WH ('Geophila') specifically to help parse and correct herbarium locality data; to manage the imprecise location of many botanical records specified only as regions, forest reserves and similar, within a range of coordinates (a bounding box); and to perform the analyses and map the results respecting this imprecision as far as possible.

Flora region data was extracted from published volumes of *Flora of Tropical East Africa* (Polhill et al., 1952-2016); *Flora of Somalia* (Thulin, 1993-2009); and the *Flora of Ethiopia and Eritrea* (Hedberg et al., 2009). For the analysis presented in Fig. 1, no data have been added to Flora regions cited for any species, even though many new collections have been made since their publication. However, grid analyses presented in Figs. 2A-C which fully include a single flora region include those records for that cell. In addition to the summary of data sources in Marshall et al. (2016), recent RBS efforts since 2016 have added 11,607 (8643 vouchered) new botanical records from Uganda and Tanzania.

Species are included for only botanical records data that are not flagged as problematic (for example, for herbarium data, some problematic records have no textual locality data and the given coordinates are probably a token country or region mid-point; or the species name may be associated with uncertainty). Botanical records relating specifically to botanic or similar gardens are also excluded. For gridded analyses, species lists are compiled for each grid cell only for botanical records whose coordinate bounding box fits entirely within a grid cell. Point records exactly on the southern or western boundary of a cell are included in that cell.

GHI is calculated using the following formula, where N_{BK}, N_{GD}, N_{BU} and N_{GN} are the number of Black, Gold, Blue and Green Star species in a grid cell or sample region, and W_{BK}, W_{GD} and W_{BU} are the respective weights (27, 9, 3).

$$GHI = 100 * (N_{BK} \times W_{BK} + N_{GD} \times W_{GD} + N_{BU} \times W_{BU})/(N_{BK} + N_{GD} + N_{BU} + N_{GN})$$

A subset of Green Star species (GX) are invisible in this GHI calculation. GX species are species whose distributions are largely if not entirely anthropogenic in tropical Africa (species like cassava or *Ixora coccinea* L.*)*, and incidentally are not regularly collected by botanists. These records were excluded from GHI calculations so that the GHI of a cell would not be unfairly brought down by the presence of a city, farm, or botanic garden (for example) in the cell.

Results and Discussion

Scaling bioquality in East Africa

The coarser the resolution of an analysis, the more botanical records are available for the analysis as a whole, and the more complete is the dataset for each location (the representation of species in the place). This is because the locality information given for some records, especially older records, is too imprecise to locate the record within fine scale sample grids or polygons; an extreme example would be 'West Afrika'. Whilst many country level records are available, we start here at the slightly more meaningful Flora region level, progressing through increasingly fine scale resolutions, and culminating with single RBS sample points. To facilitate comparison between maps at these different scales, the continuum of GHI scores in the results are classified into GHI bands, and mapped in all cases with the colours shown in Table 2. Table 2 also includes the proportion of records within each GHI band in the one degree analysis (Fig. 2A) and subsequent tenth degree analysis (a small section of which

is shown in Fig. 2C). The very hottest spots, the top 7% GHIs of all 1 degree cells, are the yellow records in class 6 with GHIs above 450.

GHI scores calculated from botanical records published in the Floras of East Africa, Somalia, and Ethiopia & Eritrea (see Methods) are shown in Fig. 1. The extreme hotspot scores in northern and eastern Somalia — the highest GHIs at any scale in Africa — are mapped in yellow (Table 2). A gradient towards lower bioquality in the heavily populated Flora regions around Lake Victoria in the west (score 146) is apparent. Flora regions were not designed to be phytochorological units, and although these Flora region scores do reveal a broad trend of bioquality across East Africa, they disguise many important trends at finer scales.

Bioquality scores across tropical Africa at one degree square resolution are shown in Fig. 2A. These scores have been updated since they were published in Marshall et al. (2016), following the latest data additions and nomenclatural changes. Trends in bioquality (which remain consistent) were discussed in that publication. A part of this one degree grid, zoomed in to East Africa, is shown in Fig. 2B. GHI values are shown inside the circles representing one degree squares. Several degree square cells have lower GHIs than the Flora region that contains them as a whole; conversely, some class 6 (very hot) hotspot cells are seen further inland than is suggested by the Flora region pattern. Although many degree square cell boundaries straddle more than one flora region, the Flora region map can now be seen to have been disguising a stretch of hotter cells in north eastern Kenya by the Somali border; and other finer scale hotspots are also revealed. For instance, the Eastern Arc biodiversity hotspot (Lovett, 1998) is now visible as the yellow cells three degrees inland from the coast at 8 degrees south in Fig. 2B). The high GHI of these Eastern Arc mountains is diluted in the Flora regions (T3, T6) by other more lowland, lower bioquality patches.

GHI scores are calculated at 0.1 degree resolution (c. 11 x 11 km squares) within two one degree squares in Fig. 2C. There are fewer eligible botanical records for the 0.1 degree square analysis, and only cells with more than 40 species are presented. Fig. 2C reveals that the GHI scores for those one degree cells, whilst still high for Africa, are an underestimate for some of the included areas. Many globally common species have not been captured in all the smaller cells; and the one degree average GHI scores are brought down by a few low ranking areas. This pattern can be seen across tropical Africa as a whole: a lower proportion of 0.1 degree resolution ('fine scale') cells have middle ranking GHIs, and a higher proportion take extreme values, compared with the one degree ('broad scale') cells (Table 2). The lowest ranking coldspot scores increased from 18% to 27% of the area of the whole tropical African grid at the fine scale; 10% of the area at 0.1 degree square resolution is seen to hold the hottest hotspot scores compared to 7% of the area at one degree resolution.

Below 0.1 degree resolution, gridded analyses become increasingly fragmented, as more and more botanical records from herbaria at least are too imprecisely located to fit in any one cell, and such lists include too few species to analyse. It is, however, possible to analyse checklists of areas smaller this. Even checklist data for small areas may be omitted from square grid summaries if they straddle two or more adjacent cells.

The two cells in Fig. 2C include the area of the Kenyan coastal forests surveyed by Hawthorne (1993), and subsequently and more completely along the Kenyan coast by Robertson & Luke (1993) and others. These coastal surveys all concluded that the coastal fragments were a priority for conservation efforts, due to the number of globally rare species, but what couldn't be answered with any precision at the time was how these values

compared with other patches of vegetation continent-wide — how much of a priority are they in terms of their rare plants? How many globally rare species counted as a lot? The GHI scores of the checklists from six of those checklists can now be calculated, and are shown in Fig. 3. All are almost as high as or higher than the GHI suggested by the 0.1 degree grid cells that include them (the gridded data includes the records from the checklist area if the latter is fully included).

The Arabuko-Sokoke Forest Reserve (F.R.) achieves an overall score of 322, spreading across several 0.1 degree grid cells that score below and above this value. The lowest of all the checklist scores (170) is for the forest adjacent to Arabuko-Sokoke F.R. within the 18 hectare Mida-Gede archaeological site, a Swahili settlement abandoned in 17th century and now partly late secondary forest, which may account for its low GHI score compared to other coastal forests. The kaya (and other coastal forests) have internationally high and important GHI scores, probably due to their long-term protection as forests sacred to the Mijikenda. These in turn contribute to the high bioquality seen at broader scales. On the other hand, we also see that some patches of the drier forests and bushland behind the coast, for instance around Kilibasi hill (Longitude 38.9 E, Latitude 3.95 S, in the cell with GHI 460 in Figure 2C; and to the north east of this, the cell with GHI 512, including the Taru and Kilisa Hills) have even higher GHI than the kaya forests.

These coastal samples were intensively sampled because they were clearly rich in rare species. In many regions relatively little botanical sampling has been conducted in very secondary vegetation, so it is likely most coarse scale analyses tend to overestimate the prevailing bioquality of the area in question. RBS of some low bioquality areas is, however, recommended, to give better context for any high bioquality samples amongst them.

The smallest area it is practical to zoom into and define a bioquality score for is the area required to hold 40 or more species, typically between 0.1 and 1 ha. At this scale, RBS or equivalent plot data is required, but very few areas have such data available. However, this is the scale at which the bulldozer operator and farmer make decisions every day.

The GHI scores of individual RBS sample points along a transect in Northern Tanzania are presented in Fig. 4. The samples were conducted as part of the botanic baseline surveying effort carried out by the authors for RSK on behalf of the East African Crude Oil Pipeline as part of their Environmental and Social Impact Assessment (Hawthorne et al., 2018) — a corridor 100m wide and 1450 km long. Each sample was conducted in a single well-defined vegetation type, and the sampling design was stratified such that areas of natural vegetation, deemed most likely to have the highest bioquality vegetation were most highly sampled. It is assumed that the intervening areas have lower bioquality than the sampled points, as much of it is fallow or bushland that is homogenous for tens of kilometres in all directions. Even the majority of the samples themselves were low bioquality, as the pipeline route had already been aligned to avoid forests and other apparently primary vegetation. RBS sampling revealed two zones where the pipe route alignment could be improved. Firstly, to the west, a newly identified location of Itigi-like thicket vegetation was identified around 50 km north-east of the furthest previously known and mapped extent of Itigi thicket, and is visible as the large red circles and high GHI scores in Fig. 4. WWF recognise White's (1983) Itigi Thicket vegetation formation as the very small (3000 square miles) and highly endangered Itigi-Sumbu thicket Ecoregion (WWF, 2019), with 50% of the Tanzanian portion already cleared (Kideghesho, 2001), and as much as 71% of the Zambian portion cleared (Almond, 2000). Secondly, to the east, back in the realm of

the East African Coastal Forests, a hotspot is identified in an area of limestone near Tanga, where the proposed pipeline would have crossed the Sigi River.

There is much to learn about fine scale bioquality patterns, even within a single hectare. An early and incomplete attempt to achieve this (Chua et al., 1996) was conducted in and around the Pasoh 50 hectare plot in Malaysia. This study, which remains incompletely published, showed that forest regenerated several decades after logging with the Malayan uniform system, apparently well, had significantly lower GHI scores than the less secondary forest of the adjacent Pasoh 50 hectare plot itself. Unfortunately only the tree community was enumerated, and without records of non-trees GHI scores *per se* could not be calculated for a strict comparison with other sites.

An advantage of the GHI is that it may be calculated for an area of any size. The utility of fine scale bioquality mapping throughout a landscape can be readily appreciated, for example when considering the placement of new infrastructure. At coarser scales, for example GHI by degree square, flora region, county level, or even country, the utility may be less transparent. The value of such coarse scores, calculated over a large scope such as tropical Africa, is that they help prioritise regions to be subsampled more intensely, to resolve the finer mosaic of hotter and colder areas that will surely be uncovered within. Even without these finer surveys, coarse scale results are sufficient in the initial stages of planning to zone parts of the country, for instance forest reserves, which may be priorities for conservation or more suitable for timber production, and to help managers allocate conservation budgets accordingly. Ultimately, for areas where no fine scale sample data are available (the very great majority of earth's surface), coarse scale data are better than nothing.

Other strengths of the RBS and bioquality approach

As range size is measured globally, and Star weightings are standardised, the bioquality hotspots in Africa can be compared with bioquality hotspot patterns derived from RBS surveys in other parts of the world. For example, the mountain forest of Tobago and northern Trinidad score similarly to those in West and Central Africa, and the lowland forest is lower in both cases (Baksh-Comeau et al., 2016; Marshall et al., 2016). The main hotspot in Japan is revealed as lying in the southern Ryuku islands (Nakamura, 2013), where scores of > 300 are comparable to the local hotspots associated with mountain peaks in Trinidad and Tobago.

In addition to bioquality scoring, RBS data can be used to classify vegetation types according to overall species content, useful for example when modelling the effects of climate change (Maharaj, 2011). We are currently using RBS data alongside herbarium data to update the chorological regions of Africa at a finer scale than in previous versions (e.g. White, 1983). The revised botanical regions of Africa will be useful, for instance, in the facilitation of digital field guide production, allowing reasonable shortlists of species to be suggested for previously under collected areas (Marshall et al., in prep.).

In areas where people depend greatly on their local vegetation for medicine and food, simple protocols have been developed to allow essential ethnobotanic data like local name, brief usage or issues of local availability to be collected alongside the regular RBS effort (Hawthorne & Marshall, 2016). Such data are useful for informing management decisions at the local scale, especially as global bioquality and local use value do not necessarily align (Marshall & Hawthorne, 2012). Field guides including local names, species illustrations and Latin names

can facilitate communication about plants between managers, scientists and local residents, democratising botany and improving conservation outcomes (Marshall & Hawthorne, 2013). A field guide produced for the flora under study and illustrated in any format will increase users' plant identification accuracy; in fact a novice non-botanist may achieve accuracy rates almost as high as experienced botanists when provided with suitable guide material, even in species rich tropical rainforest (Hawthorne et al., 2014). RBS field work provides ample opportunity for the next generation of botanists to hone their skills.

Conclusions

Past hotspot analyses have shown global hotspots at broad scale, or local hotspots referring only to rarity within the context of a sampled region. The RBS and bioquality approach shows how global rarity can be integrated into maps at any scale, to define local to regional hotspots based on the overall global rarity of the species in the community. Strategic and day-to-day decisions about how to defend against biodiversity loss are facilitated.

Although fine scale and intensive RBS or equivalent data (checklists for small areas) add a lot to the broad and speckled picture painted by traditional herbarium databases, networks of very local, very complete inventories for particular landscape units and vegetation types are the exception. An RBS can provide current data for consistent and standardised EIAs and for vegetation and biodiversity monitoring, filling a different role to measured plots and taxonomic-style herbarium collections. They could vastly improve the standard of botanic databases in the coming decades, complementary to herbaria, measured plots and satellite information and fit for use with the intense monitoring and subtle fine scale land management decisions that will become ever more important.

There is a need to increase the intensity or coverage of National botanical surveys, funded centrally and integrated with university teaching and herbarium strategies, Environmental Protection Agencies and Environmental impact assessors. These will cost hundreds of thousands or millions of dollars for all but the smallest countries, but revenue could be raised from developers commissioning surveys in areas to be exploited, international development banks, and environmental NGOs. Herbaria, universities and/or botanic gardens, could coordinate the effort and maintain online databases with high taxonomic standards, IUCN threat listings and Star ratings. This would revitalise field botany and benefit biodiversity management.

Although RBS does generate a steady flow of orthodox fertile herbarium specimens with duplicates, for the more interesting species, the majority of vouchers, thousands per month, are unicate, sterile specimens for which herbaria usually have no space. Any commitment to a national RBS should include a parallel RBS voucher herbarium, where the dried specimens are archived and well-organised, but can remain unmounted, but labelled in their original pressing newspaper. In an era when DNA and other biochemical profiling is ever easier, these stockpiles of less precious specimens will be become an important resource. They are also excellent reference material for teaching, identification of sterile plants, and the production of field guides.

RBS and bioquality are now well established in tropical Africa, and sizable studies have been carried out in the Neotropics (Trinidad & Tobago; Maule region Chile; Brazilian cerrado). Thus far RBS has only been tried twice

in Asia in an extensive way: in Japan (Nakamura, 2013; ongoing); briefly in Pasoh Forest Reserve (Chua et al, 1996), and at a few trial samples in Brunei (Kuala Belalong Field Studies Centre). We hope to encourage biodiversity managers of Asian landscapes to try it, and to help develop the global coverage of globally-standardised bioquality scores.

References

African Plants Database (APD) (version 3.4.0). Conservatoire et Jardin botaniques de la Ville de Genève and South African National Biodiversity Institute, Pretoria, Retrieved June 2014, from http://www.ville-ge.ch/musinfo/bd/cjb/africa/.

Almond, S. (2000). *Itigi thicket monitoring using Landsat™ Imagery*. MSc. Remote sensing dissertation. University College, London.

Baksh-Comeau, Y.S., Maharaj, S.S., Adams, C.D., Harris, S.A., Filer, D.L. & Hawthorne, W.D. (2016). An annotated checklist of the vascular plants of Trinidad and Tobago with analysis of vegetation types and botanical "hotspots." *Phytotaxa* 250: 1–431.

Barthlott, W., Mutke, J., Rafiqpoor, D., Kier, G., and Kreft, H. (2005). Global Centers of Vascular Plant Diversity. *Nova Acta Leopoldina* NF 92 342: 61–83.

Brooks, T.M., Mittermeier, R.A., Fonseca, G.A.B. da, Gerlach, J., Hoffmann, M., Lamoreux, J.F., Mittermeier, C.G., Pilgrim, J.D. & Rodrigues, A.S.L. (2006). Global Biodiversity Conservation Priorities. *Science* 313: 58–61.

Burgess, N.D., Clarke, G.P. & Rodgers, W.A. (1998). Coastal forests of eastern Africa: status, endemism patterns and their potential causes. *Biological Journal of the Linnean Society* 64: 337–367.

Burgess, N.D., Mwasumbi, L.B., Hawthorne, W.D., Dickinson, A. & Doggett, R.A. (1992). Preliminary assessment of the distribution, status and biological importance of the coastal forests in Tanzania. *Biological Conservation* 62: 205–218.

Chua L.S.L, Hawthorne, W.D., Saw L.G. & Quah E.S. (1996). Biodiversity database and assessment of logging impacts. In L.S. See, D.Y. May, I.D. Gauld & J. Bishop (eds), *Conservation, Management and Development of Forest Resources*, pp. 30–40. London: International Institute for Environment and Development.

Department of Plant Sciences, University of Oxford. (2013). *Botanical Research and Herbarium Management System: Documentation — BRAHMS Online*. http://herbaria.plants.ox.ac.uk/bol/brahms (accessed 08 May 2019).

Friis, I. & Vollesen, K. (1989). Notes on the vegetation of southernmost Somalia, with some additions to the flora. *Willdenowia* Bd. 18, H. 2: 455–477.

Govaerts, R. et al. [multiple collaborators] (2019). *World Checklist of Selected Plant Families*. Facilitated by the Royal Botanic Gardens, Kew. http://apps.kew.org/wcsp (accessed 2014–2019).

Hall, J.B. & Swaine, M.D. (1981). *Distribution and ecology of vascular plants in a tropical rain forest: forest vegetation in Ghana.* Geobotany 1. Junk, The Hague

Hawthorne, W.D. (1993). East African coastal forest botany. In J.C. Lovett & S.K. Wasser (eds), *Biogeography and Ecology of the Rain Forests of Eastern Africa*, pp. 57–99. Cambridge University Press.

Hawthorne, W.D. (1996). Holes and the sums of parts in Ghanaian forest: regeneration, scale and sustainable use. *Proceedings of the Royal Society of Edinburgh Section B: Biological Sciences* 104: 75–176.

Hawthorne, W.D. & M. Abu-Juam. (1995). *Forest Protection in Ghana*. Gland and Cambridge: IUCN/ODA/Forest Department Republic of Ghana,.

Hawthorne, W.D., Grut, M. & Abu-Juam, M. (1998). *Forest production and biodiversity conservation in Ghana, and proposed international support of biodiversity conservation.* CSERGE working paper, GEC 98–18.

Hawthorne, W.D., Sheil, D., Agyeman, V.K., Abu Juam, M. & Marshall, C.A.M. (2012). Logging scars in Ghanaian high forest: towards improved models for sustainable production. *Forest Ecology and Management* 271: 27–36.

Hawthorne, W.D., Cable, S. & Marshall, C.A.M. (2014). Empirical Trials of Plant Field Guides. *Conservation Biology* 28: 654–662.

Hawthorne W.D. & Marshall, C.A.M. (2016). *A manual for Rapid Botanic Survey (RBS) and measurement of vegetation bioquality 2016*. http://herbaria.plants.ox.ac.uk/bol/oxford/Survey.

Hawthorne, W. D., Ndangalasi, H., Mbago, F., Selemani, H., Kalughasha, J. & Marshall, C.A.M. (2018). *East Africa Crude Oil Pipeline (EACOP) Tanzania Export Pipeline Environmental and Social Impact Assessment —* Appendix A1: Biodiversity Baseline Report (Botany Component).

Howard, P.C., Davenport, T.R.B., Kigenyi, F.W., Viskanic, P., Baltzer, M.C., Dickinson, C.J., Lwanga, J., Matthews, R.A. & Mupada, E. (2000). Protected area planning in the tropics: Uganda's national system of forest nature reserves. *Conservation Biology* 14: 858–875.

IUCN (2019). *The IUCN red list of threatened species. Version 2019-1*. http://www.iucnredlist.org.uk Downloaded on 21 March 2019.

Kideghesho, J.R. (2001). The status of wildlife habitats in Tanzania and its implications to biodiversity. *Tanzania Wildlife* 21: 9–17.

Kier, G. & Barthlott, W. (2001). Measuring and mapping endemism and species richness: A new methodological approach and its application on the flora of Africa. *Biodiversity & Conservation* 10: 1513–1529.

Lindsell, J., Agyei, R., Bosu, D., Decher, J., Hawthorne, W., Marshall, C., Ofori-Boateng, C. & Rödel, M.-O. (2019). *The biodiversity of Atewa forest research report*. Accra: A Rocha Ghana.

Lovett, J.C. (1988). Endemism and affinities of the Tanzanian montane forest flora. *Monographs in Systematic Botany from the Missouri Botanical Garden* 25: 591–598.

Lovett, J.C. (1993). Eastern Arc moist forest flora. In J.C. Lovett & S.K. Wasser (eds), *Biogeography and ecology of the rain forests of eastern Africa,* pp. 33–55. Cambridge University Press.

Lovett, J.C. (1998). Importance of the Eastern Arc Mountains for vascular plants. *Journal of East African Natural History* 87: 59–74.

Lovett, J.C., Bridson, D.M. & Thomas, D.W. (1988). A preliminary list of the moist forest angiosperm flora of the Mwanihana Forest Reserve, Tanzania. *Annals of the Missouri Botanical Garden* 75: 874–888.

Lovett, J.C., Hatton, J., Mwasumbi, L. B. & Gerstle, J. (1997). Assessment of the impact of the lower Kihansi hydropower project on the forests of Kihansi Gorge, Tanzania. *Biodiversity and Conservation* 6: 915–933.

Maharaj, S.S. (2011). *The impact of climate change on the small island developing states of the Caribbean.* D.Phil. Thesis, University of Oxford.

Marshall, C.A.M. & Hawthorne, W.D. (2012). Regeneration ecology of the useful flora of the Putu Range rainforest, Liberia. *Economic Botany* 66: 398–412.

Marshall, C.A.M. & W.D. Hawthorne (2013). *Important plants of northern Nimba County, Liberia. A guide to the most useful, rare or ecologically important species, with Mano names and uses.* Oxford Forestry Institute.

Marshall, C.A.M., Wieringa, J.J. & Hawthorne, W.D. (2016). Bioquality hotspots in the tropical African flora. *Current Biology* 26: 3214–3219.

Meerts, P. (2016). An annotated checklist to the trees and shrubs of the Upper Katanga (D.R. Congo). *Phytotaxa* 258: 201–250.

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

Nakamura, N. (2013). *Dissecting the Japanese hotspot.* D. Phil. Thesis, University of Oxford.

Olson, D.M. & Dinerstein, E. (2002). The Global 200: Priority Ecoregions for Global Conservation. *Annals of the Missouri Botanical Garden 89*: 199.

Polhill, R., Beentje, H., et al. (eds) (1948–2012). *Flora of Tropical East Africa.* Various volumes. Royal Botanic Gardens, Kew & Rotterdam: Balkema.

Robertson, S.A. & Luke, W.R.Q. (1993). *Kenya Coastal Forests. The report of the NMK/WWF Coast forest survey.* WWF Project 3256, Coast Forest Status, Conservation and management.

Stattersfield, A.J., Crosby, M.J., Long, A.J. & Wege, D.C. (1988). *Endemic bird areas of the world: priorities for biodiversity conservation.* Cambridge: BirdLife International.

Stropp, J., Ladle, R.J., M. Malhado, A.C., Hortal, J., Gaffuri, J., H. Temperley, W., Olav Skøien, J. & Mayaux, P. (2016). Mapping ignorance: 300 years of collecting flowering plants in Africa. *Global Ecology & Biogeography* 25: 1085–1096.

Thulin, M. (ed.) (1993–2009). *Flora of Somalia.* 4 Volumes. Royal Botanic Gardens, Kew.

Westhoff, V. & van der Maarel, E. (1978). The Braun-Blanquet approach. In: R.H. Whittaker (ed.), *Classification of plant communities*, pp. 289–312. The Hague: Junk.

White, F. (1983). *The vegetation of Africa — a descriptive memoir to accompany the Unesco/AETFAT/UNSO vegetation map of Africa*. Natural Resources Research XX. Paris, France: UNESCO.

Wikum, D.A. & Shanholtzer, G.F. (1978). Application of the Braun-Blanquet cover abundance scale for vegetation analysis in land development studies. *Environmental management* 2: 232–329.

Williams, P. (1996). Promise and problems in applying quantitative complementary areas for representing the diversity of some Neotropical plants (families Dichapetalaceae, Lecythidaceae, Caryocaraceae, Chrysobalanaceae, and Proteaceae). *Biological Journal of the Linnaean Society* 58: 125–157.

WWF (2019). Itigi-Sumbu thicket. Available at https://www.worldwildlife.org/ecoregions/at0708 and accessed 08 May 2019.

Figure captions

Fig. 1. East African Flora regions showing the GHI calculated for species listed in the flora for these regions. Colours indicate bands of GHI based on continent wide percentile. The hottest GHI hotspot at this scale is NE Somalia; a downward gradient inland is apparent.
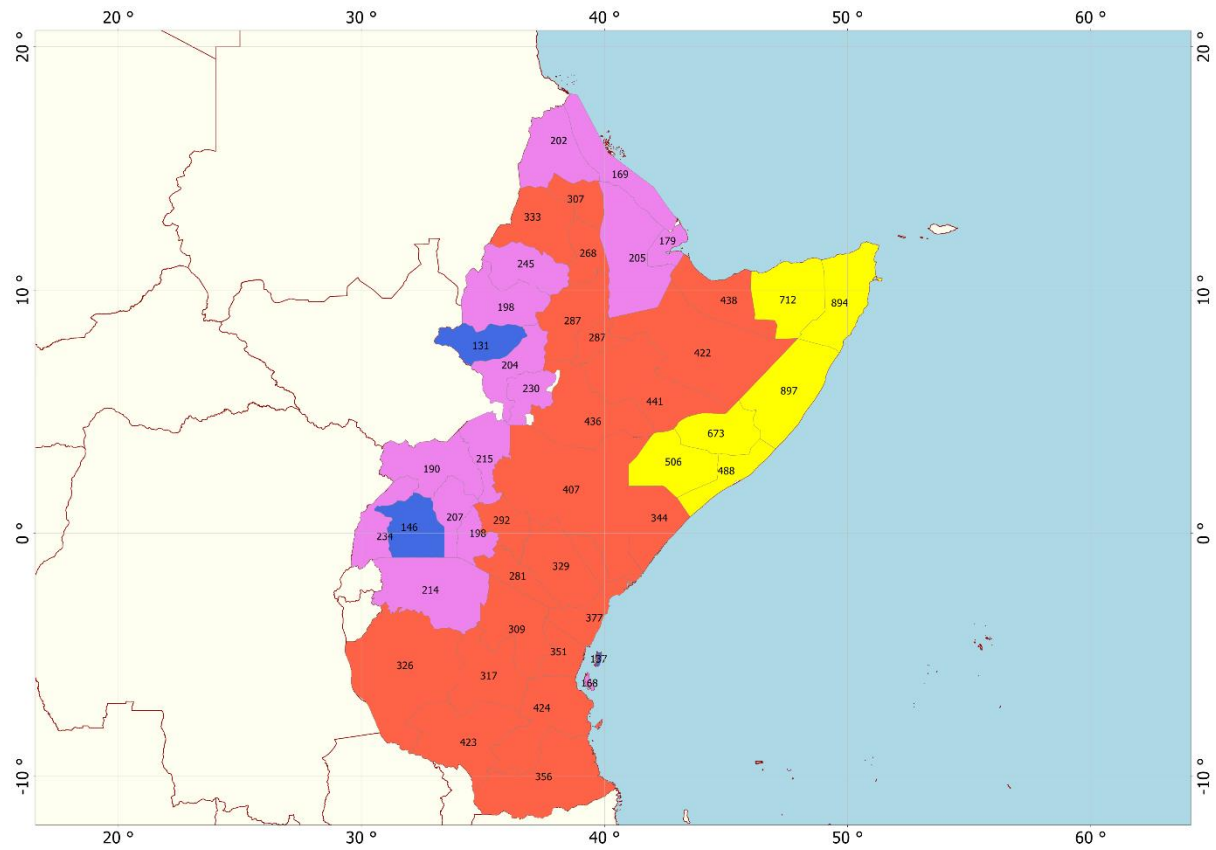
Fig. 2. A) Bioquality scores across Africa at 1-degree resolution, update from Marshall et al. (2016). Circles represent the degree squares for which adequate data are available, presented against a black background and coloured by GHI class — see colour key that applies to all maps. B) the same map zoomed in to the East African coast. C) A portion of the area of 2b, as indicated on 2b, showing two grid cells subdivided and analysed at 0.1 degree resolution. The 1 degree cell score is shown in larger orange bordered font.
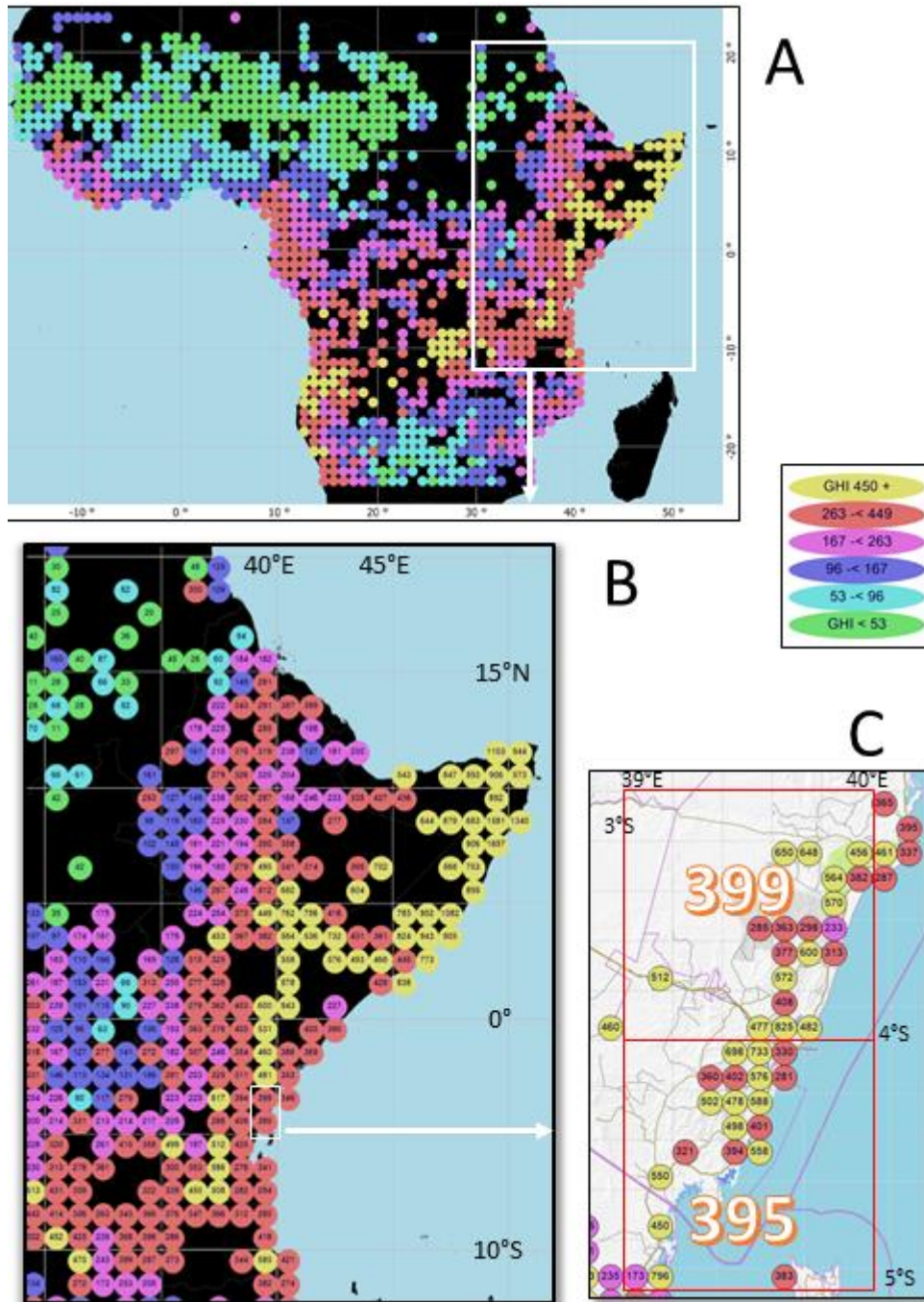
Fig. 3. *Map of the Kenyan coastal region as in Figure 2c, showing location (pale green boxes) of 6 individual forests for which the GHI values in the pale green boxes have been calculated directly from forest checklists. Each pale green box is the minimum bounding box that encloses the outline of each forest, and is much smaller than the blue outline 0.1 degree cell that includes them, except the Arabuko-Sokoke F.R. which straddles several cells and for which the forest is shown as the green polygonal region, with pale green box (GHI=322) at its centre . The GHI of each enclosing 0.1 degree cell is shown in larger orange font.*
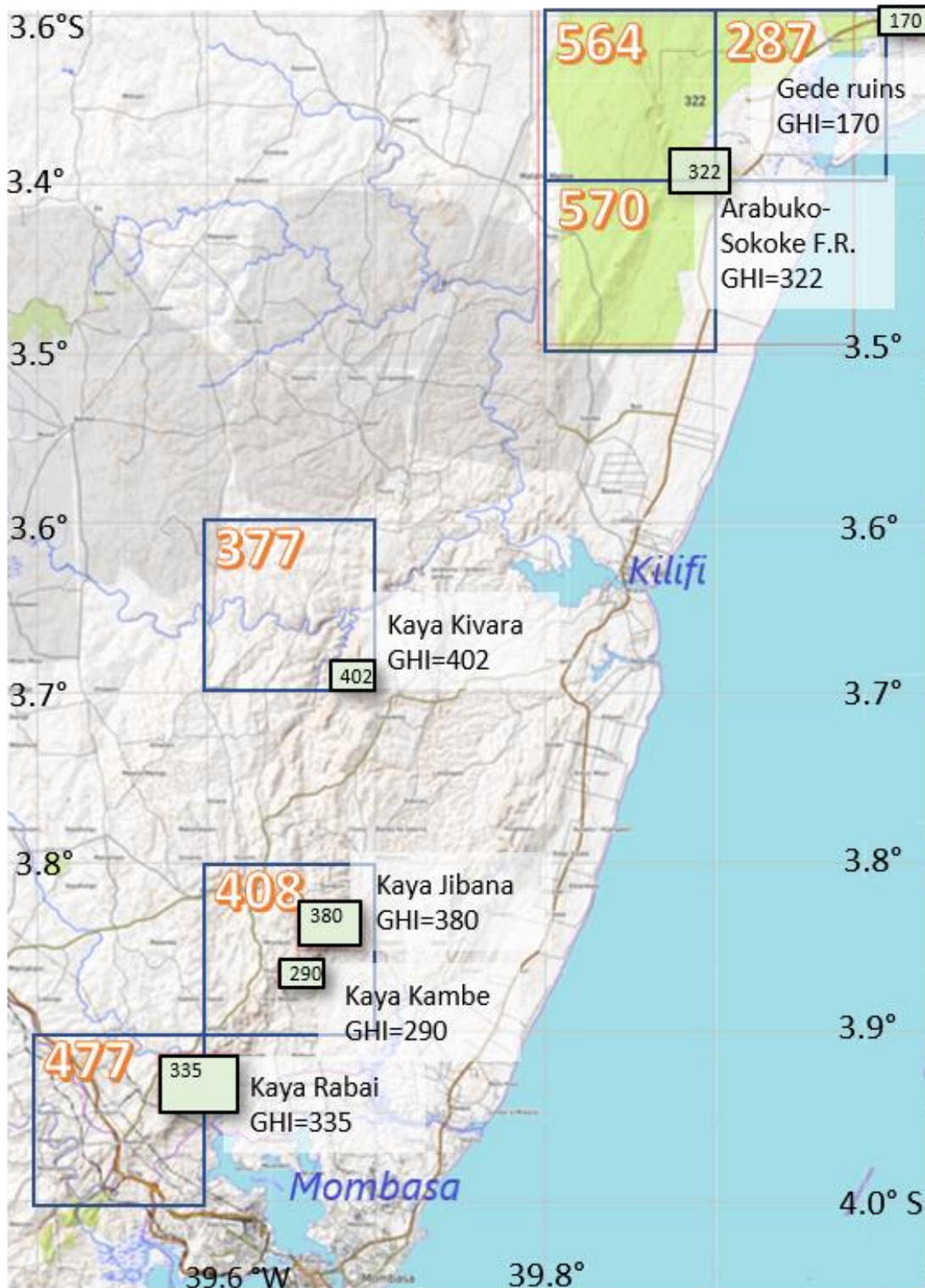
Fig. 4. RBS samples along a part of transect in northern Tanzania, from Tanga port in the east to the vicinity of Singida in the west (c 580km west to east). All samples (<1 ha) would be small dots at this scale, but have been sized in proportion to GHI. Patches of Itigi thicket vegetation have a high bioquality in the west, with GHI reaching above 400. Samples of disturbed coastal forest and thicket in the east are also highlighted. The GHI declines in the mosaic of farm, woodland and bushland in between.
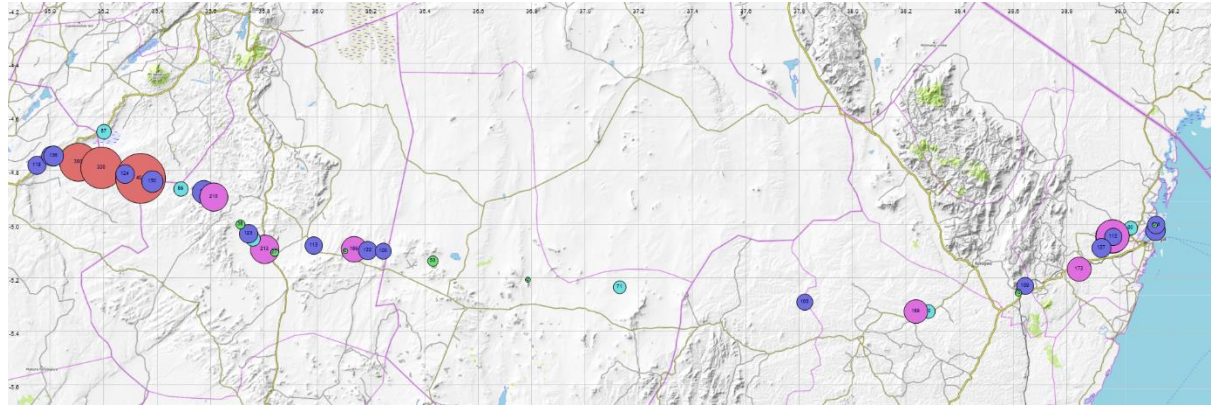
Tables

Table 1. Ideal, mean geographic ranges and standard weights for each Star

Star	Ideal mean range (tropical degree square or 100 × 100 km)	Standard Weight
Black	2.66	27
Gold	8	9
Blue	24	3
Green	72	0 for calculation of GHI

Table 2. GHI classes and ranges for the colour scheme mapped in Figs. 1-4. The number of records in each GHI class is summarised at one degree, and 0.1 degree square resolution.

GHI class	GHI range	Number of records at 1 degree square	% (n= 1263)	Colour on maps	% of 0.1 cells (n=15163)
1	0-52	228	18%	Green	27%
2	53-95	233	18%	Cyan	17%
3	96-166	233	18%	Purple	17%
4	167-262	236	19%	Pink	15%
5	263-448	240	19%	Orange-red	14%
6	449-1657	93	7%	Yellow	10 %