The effect of forest restoration on butterfly diversity and composition in the Lower Kinabatangan, Sabah, Malaysian Borneo

Rhys Davies

Degree Scheme: BSc Biological Sciences (Zoology) with a Professional Training Year

Vocational Supervisor: Prof. Benoit Goossens

Word Count: 5778

Table of Contents

Part A - Personal Reflection

From 9 September 2022 to 26 August 2023, I spent my professional training year (PTY) placement at Danau Girang Field Centre (DGFC), located in Lot 6 of the Lower Kinabatangan Wildlife Sanctuary (LKWS) in the state of Sabah, Malaysian Borneo. At any given time, DGFC is home to roughly twenty researchers, students, volunteers and local Malaysian staff. The station is very isolated and at 11pm every night, the power is turned off and you find yourself completely immersed within the fantastic sounds and absolute darkness of the Bornean rainforest.

After first hearing about DGFC whilst discussing PTY opportunities with one of my lecturers, I was immediately sold on the idea of living in the rainforest for a year and immersing myself in tropical biodiversity research. This year has truly been the opportunity of a lifetime and something I recognise that most people will never be fortunate enough to experience. I will always be incredibly grateful and indebted to the director, Dr Benoit Goossens, for affording me this truly incredible opportunity.

The range of work in which I have been able to immerse myself and develop my practical competencies in has been fantastic. I have gained experience in VHF tracking Pangolins and UHF tracking Leopard Cats in the surrounding forests and plantations, assisted in a procedure to radio tag two pangolins, bat sampling, small mammal trapping of rodents, mesocarnviore trapping of monitor lizards and civets, mist netting, pitfall trapping and butterfly surveys. There is also a heavy use of camera traps and I have thoroughly enjoyed gaining experience with all these research methods. I have also enjoyed developing my scientific communication skills through my role as chief editor of the “Jungle Times” newsletter, interviewing visitors and writing articles about the ongoing activities at the centre. A particular highlight included interviewing the British High Commissioner to Malaysia at the time, His Excellency Charles Hay.

Furthermore, not only has this year developed my practical capabilities but I feel I have undergone a lot of personal growth and development. As a naturally quiet and introverted person, living in isolation with people (and with privacy a completely alien concept) I feel much more confident in my communication and team working skills – being able to both effectively work as part of a team but also step up and lead when necessary. I have enjoyed being repeatedly pushed out of my comfort zone, especially leading activities, presenting my work and networking with the constantly revolving door of visiting researchers, students and field course groups from other universities.

This year has certainly not been without its challenges; my body has found living in a humid and tropical environment for a year, away from my temperate home of the UK, to be incredibly challenging and I have become tiresome of the persistent and numerous visits to doctors and daily medicating routines. However, for it to have enabled me to continue living in such an incredible environment has been worth it and if someone was to have informed me of the difficulties that I would have faced all year long, I still wouldn't have even considered the possibility of changing. I think this has been a testament to my resilience and perseverance, as many people I have talked to have said they would have thrown in the towel and caught the next flight home if faced with the same situation. I am very grateful for the support and encouragement provided by my peers during those particularly hard times.

Finally, I will always fondly remember the local staff, volunteers and researchers I have had the privilege of working alongside who I can safely say are some of the most amazing people I will ever meet. Taking part in the Hari Raya celebrations in the nearby village of Batu Puteh, being invited into the homes and having the chance to meet the families of many of the staff was a very touching moment. On one last note, a particular favourite memory of the year would, without question, have been my involvement in the radio tagging of a male and female Sunda pangolin and seeing a wild baby pangolin (a sight which very few people on earth will ever see) and is a memory I will forever treasure.

Part B - Scientific Report

Abstract

The island of Borneo in South East Asia is home to some of the world's oldest tropical rainforests and highest levels of biodiversity and endemism; however, since the latter twentieth century Borneo has experienced high levels of deforestation in favour of timber production and agricultural land conversion. Given the important role forests play in mitigating climate change, forest restoration efforts and the subsequent monitoring of restoration are important to observe the recovery of forests to their natural state. Butterflies are a widely used taxonomic group for biodiversity assessments due to their relatively comprehensive taxonomic documentation and ease of species identification. This study took place in the Lower Kinabatangan floodplain and aimed to examine changes in alpha diversity (species richness, abundance, and diversity indices) and community composition of butterflies by using weekly transect surveys (across seven weeks) at two different active restoration sites and restored forest plots under the new Regrow Borneo initiative, by also comparing two areas of natural (secondary) forest and two oil palm plantations. Overall, measures of alpha diversity showed the restored forest plots and monoculture plantation had the seemingly lowest diversity, while the active restoration, high conservation value (HCV) plantation and natural forest sites had comparably higher levels of diversity. There was no clear trend of directionality with restoration age however community composition of butterflies was distinct between plantation and natural forest sites, with active restoration and restored forest sites somewhere intermediary. Environmental variable analysis suggested species richness and abundance increased at sites with a lower density of trees and relatively hotter temperatures. While this study didn't identify clear trends in diversity and composition with forest restoration unlike other studies, it does provide a foundational set of data for future butterfly monitoring under the Regrow Borneo project to be expanded on.

1.0 Introduction

Tropical forests are the most diverse ecosystems on earth, providing a home to two thirds of all plant and animal species, despite only covering 7% of the planets land surface (Bierregaard et al. 1992). In particular, the dipterocarp forests of the Sundaland in south east Asia are the oldest tropical forests in the world and Pleistocene climactic events led to a high level of biodiversity and endemism (Brookfield and Byron 1990). However, this same region has experienced a twenty fold increase in human population over the last few centuries, with exploitation of natural resources such as timber and agricultural development of palm oil (*Elaeis guineesis*) fuelling economic growth (Brookfield and Byron 1990). As a result, the high degree of endemism combined with extensive logging and forest conversion has left the biodiversity of the Sundaland some of the most imperilled (Fisher et al. 2011).

The island of Borneo is an intensively logged part of the Sundaland, losing forest cover at a rate nearly twice that of other tropical forest regions since commercial logging began in the 1970's (Achard et al. 2002). Since 1973, forest cover is estimated to have reduced by over 30% from an initial 75% and less than 28% of the remaining forest is intact (Gaveau et al. 2014). The lucrative values of timber and oil palm has fuelled land conversion (Fisher et al. 2011) with an estimated 10% of Borneo covered in oil palm and timber plantations in 2010 (Gaveau et al. 2014). The agricultural expansion of palm oil is limited to low elevation areas, conflicting with lowland tropical forests and riverine floodplains (Abram et al. 2014). The Malaysian state of Sabah (Borneo) is a major producer of palm oil and has seen total forest cover, particularly in the lowlands, decline by approximately 40% since 1973, predominantly in favour of establishing plantations (Gaveau et al. 2014).

Deforestation and land conversion can have damaging effects on biodiversity and ecosystem functioning; for example, forests afford watershed protection, ground stability (Lamb et al. 2005) and the sequestration of atmospheric carbon (Gaveau et al. 2014). The impact of logging on diversity is debated, however, the consistent results are the loss of forest dependent and often endemic species with restricted habitat requirements which can't occur in degraded or converted landscapes such as palm oil (Benedick et al. 2006). To maintain biodiversity and ecosystem functioning, preserving forests is crucial (Hamer et al. 2003) but forest restoration efforts are also critical (Lamb et al. 2005). Degradation can lead to reductions in soil fertility and domination of grasses reducing the recolonization ability of woody plant species (Lamb et al. 2005). Active restoration involves the clearing of grasses and planting of fast growing pioneer species to quickly re-establish a canopy, allowing for colonisation by a wider variety of climax species (Lamb et al. 2005).

Combating climate change, forest restoration is an important tool and therefore understanding the recovery of biodiversity and ecosystem functioning is important to assess restoration efforts (Korkiatupa et al. 2023). Long term monitoring can reveal changes in animal and plant community composition and suggest if restoration methods are stagnating and require planting of climax species (Korkiatupa et al. 2023). Biodiversity assessments require the choice of study taxa, and invertebrates (particularly butterflies) are commonly used (Stork et al. 2003). A prominent example is the UK Butterfly Monitoring Scheme which had been detecting national trends in butterfly diversity annually since 1976 (Pollard 1977) which can help inform policy decision making.

There are approximately 20,000 species of butterfly worldwide, of which 90% are found in the tropics (Bonebrake et al. 2010). Butterflies are a particularly useful indicator taxa because of their well quantified taxonomy, identifiability to species level, sensitivities to environmental and microclimate changes and result replicability (Bonebrake et al. 2010, Stork et al. 2003). Butterflies have important ecosystems roles as pollinators and with specific plant species preferences for feeding and hosting their larval stage, butterfly diversity can be related to the diversity and health of their habitats (Sparrow et al. 1994, Kremen et al. 1992). Butterfly studies, particularly in the tropics have typically focused on comparing areas of pristine and degraded forest (Bonebrake et al. 2010), with few studies focusing on restoration (Korkiatupa et al. 2023) and monitoring schemes aren't as prevalent as temperate areas due to resource and cost constraints despite the significantly higher biodiversity and ecosystem value (Bonebrake et al. 2010).

Regrow Borneo is a recent initiative, started in 2019, in the Lower Kinabatangan, Sabah, which aims to restore areas of previously degraded forest, improve the scientific understanding of forest restoration while also supporting local livelihoods (Regrow Borneo, 2022). The various active restoration and restored forest sites the project restores and studies, provide an opportunity to investigate the effect of forest restoration on butterfly species diversity (which isn't one of the taxa assessed under their biodiversity surveys) in the lowland tropical forest of the Kinabatangan. Furthermore, this also presents the opportunity to begin a longer term monitoring program for butterfly diversity for which this study has collected the initial data. Aside from providing initial data for future comparisons, the main aims of this study were to investigate:

1) How does butterfly diversity compare between the Regrow sites (active and restored) and the control sites (natural forest and oil palm plantation)?

2) Is there a directional trend of butterfly diversity along the forest restoration gradient of active restoration to restored forest to natural forest?

3) Do environmental characteristics predict any trends of butterfly species richness and abundance which could then be expected as forest restoration occurs?

2.0 Materials and Methods

2.1 Study Sites

This study took place in the Lower Kinabatangan floodplain, Eastern Sabah, within the landscape surrounding the Danau Girang Field Centre (DGFC) (05° 24' 48'' N, 118° 02' 16'' E)(figure 1). The area is characterised by a wet and humid tropical climate, average annual precipitation of 3000mm and monthly temperatures of 21-34 °C (Ancrenaz et al. 2004). The region contains a mixture of forest types, particularly freshwater swamps and lowland dipterocarp forests, much of which are susceptible to varying degrees of flooding (Anrenaz et al. 2004). The Kinabatangan has experienced intense logging activity since the mid twentieth century, leaving heavily exploited and highly fragmented forests, and over half of the floodplain is occupied by palm oil (Abram et al. 2014). In 2005, the Sabah State Government established the Lower Kinabatangan Wildlife Sanctuary (LKWS); a series of ten lots along the Kinabatangan River, totalling approximately 27960 ha of protected (but degraded) forest, acting as a wildlife corridor and refuge connecting a fragmented network of forest reserves (Abram et al. 2014).

This study focused on the Pin Supu Forest Reserve (PSFR) (figure 1A), owned by the Sabah Forestry Department (SFD), covering an area of selectively logged forest approximately 4696 ha

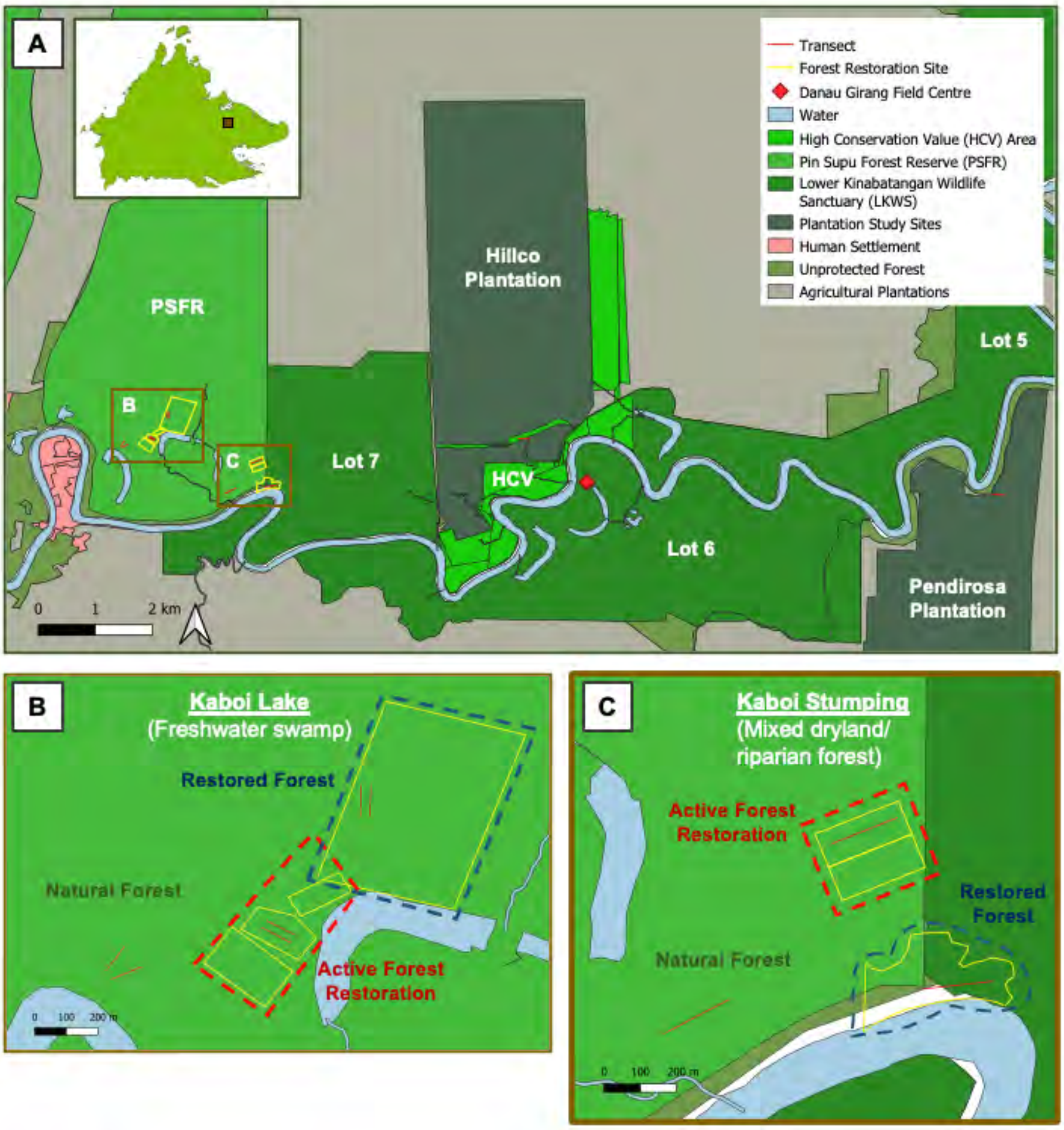


Figure 1. *Map of study area*. **A)** Overview of the Lower Kinabatangan floodplain area with the four study areas where transects were located: Kaboi Lake **(B)**, Kaboi Stumping **(C)**, Hillco plantation HCV area and Pendirosa plantation. Throughout this report Hillco and Pendirosa are referred to by their environment type: high conservation value (HCV) and monoculture (MNC) respectively. **(B)** and **(C)** show the active restoration, restored forest and natural forest study sites within each forest area. (QGIS, version 3.26.3)

which provides a wildlife corridor between Lots 7 and 8 of the LKWS. Koperasi Pelancongan (KOPEL), is a community based organisation which started restoring forest in the PSFR in 1999. Through ecotourism activities, local employment is generated and KOPEL has been managing forest restoration within the PSFR for more than 20 years. The Regrow Borneo initiative was launched in 2019, between KOPEL and DGFC, as a collaborative forest restoration effort to both benefit forest biodiversity and the livelihoods of local people.

Within the PSFR there were two study areas of focus: a freshwater swamp site “Kaboi Lake” (KL) and a mixed area of dry and riparian forest “Kaboi Stumping” (KS) (figure 1B and 1C) . Each area contains three study sites: an active site of forest restoration, a previously restored forest site and an area of natural (secondary) forest. The active sites of forest restoration are characterised by an open canopy and sparse tree cover (figure 2A), ground vegetation is routinely cleared by workers to help saplings grow. Seeds are collected from the forest and saplings are raised in the local nurseries to be planted. Restored forest sites are previously planted sites, which are no longer actively maintained and are characterised by a closed canopy. KL is an area of freshwater swamp, often prone to flooding (figure 2B), the active site has been managed since 2020 and the restored forest was originally planted in 2003. KS is an area of riparian forest, which had previously been cleared and used as a stumping ground for commercial logging activities. The active site has been managed since 2021 and the restored forest was replanted in 2007 (figure 2C). Both KL and KS have a transect located within their natural forest (figure 2D).

In addition to the two natural forest sites, two Oil Palm Plantation sites were also included in this study as control sites. The Hillco plantation (figure 2B) contains an area of forest referred to as High Conservation Value (HCV), creating wildlife corridors and areas of high biodiversity within the plantation landscape (Kwatrina et al. 2018), while the Pendirosa plantation (figure 2A) provided an example of a monoculture plantation.

2.1 Data Collection

Each site contained a 200m transect which was surveyed once a week, for seven consecutive weeks between 27th March and 11th May 2023, using a modified pollard method (Pollard 1977, Sparrow et al. 1994). The KL active site was too small of a plot for a single 200m transect, and instead two 100m transects in parallel were used for each of the three sites at KL. Data collection took place between 9am and 12pm each day, transects were walked at a constant pace (seven minutes per 100m) to ensure equal sampling for each site. Sampling only took place on days with

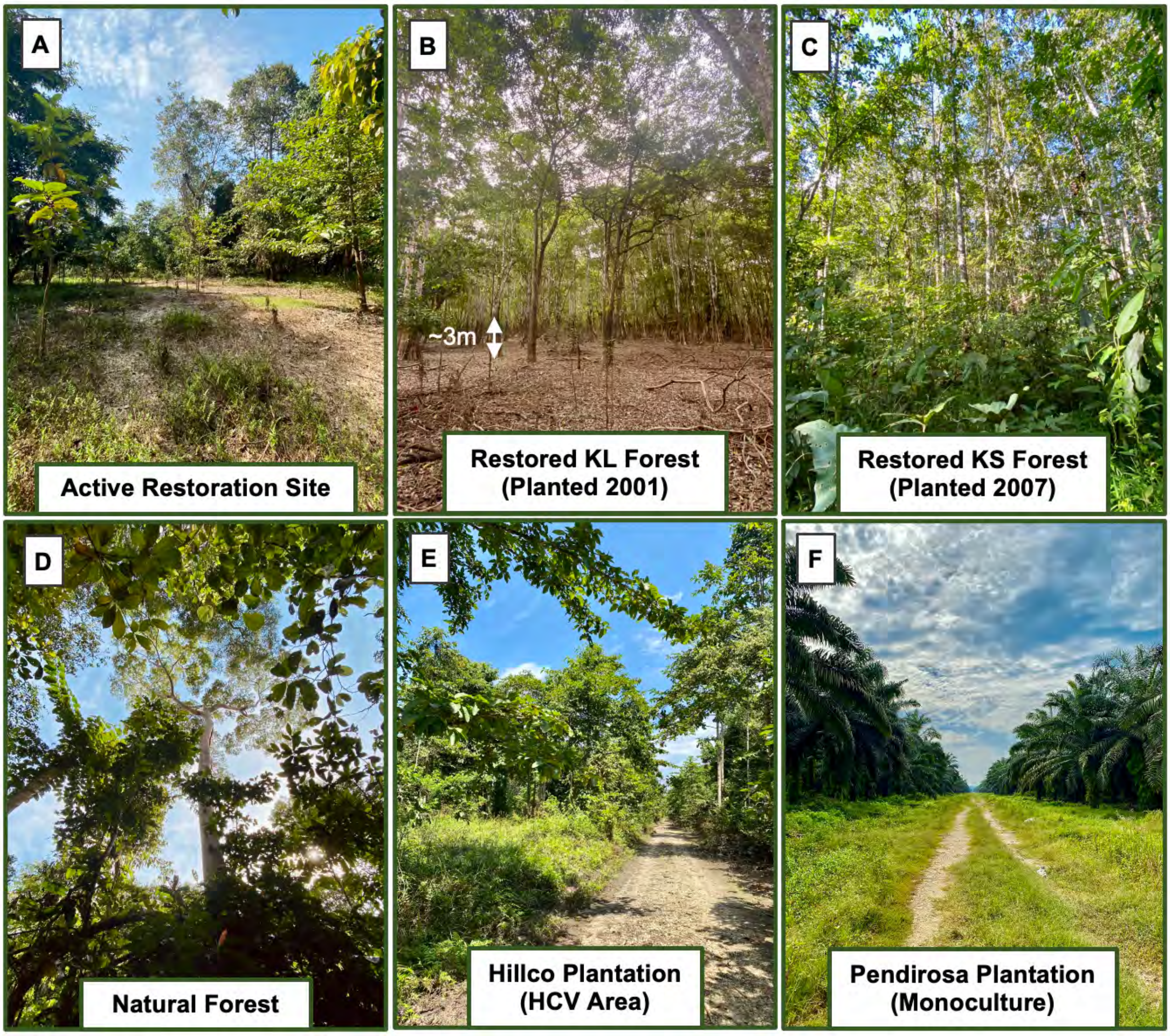


Figure 2. *Butterfly study site types.* **A)** Site of ongoing forest restoration; at each active site ground vegetation is routinely cleared to aid sapling growth. **B)** Kaboi Lake (KL) forest restoration site (planted 2003), a freshwater swamp site, water levels can reach approximately 3 metres high during periods of flooding. **C)** Kaboi stumping, riparian restored forest (planted 2007). **D)** Natural (secondary) forest site, forests across the Kinabatangan were repeatedly logged throughout the last century. **E)** Area of high conservation value (HCV), Hillco plantation. **F)** Monoculture palm oil, Pendirosa plantation.

at least fine weather; if a sampling day was cancelled due to unsuitable weather conditions or flooding, sampling was rolled over to the next available day.

Individuals within five metres either side of the transect were counted and individuals behind the sampling team were not counted to avoid replicates. Individuals were either visually identified, caught with a butterfly net and identified in the field or photographed for post sampling identification using a suitable field guide (Otsuka 2001). Butterflies of the genus “Tanaecia” and “Euthalia” require lab dissection of the genitalia to be identified to species level (Scriven et al. 2016) so individuals were recorded under a “Euthalia/Tanaecia” genus grouping. Butterflies of the genus “Eurema” proved difficult to reliably identify to species level and remained grouped to generic level.

These two genus groupings were treated as species for the data analysis.

2.2 Habitat Characterisation

The environmental characteristics of each transect was characterised by a habitat assessment. At the beginning and end of each transect sampling, temperature (°C) and humidity (%rh) were recorded using EL-USB-2 data loggers (Lascar Electronics Ltd) which were averaged together to provide each sampling effort with an overall reading. At each site: estimated percentage ground (0-2m) and lowstorey (2-5m) vegetation cover (to the nearest 5%), ground vegetation depth, ground vegetation density, canopy cover, canopy height and the density of saplings, small trees and large trees was measured.

Four habitat assessments per 200m transect were carried out (using randomly generated distances) where a central point was marked on the transect and four marker sticks were placed four metres away from the central point at North, East, South and West. This provided five measurement points per habitat assessment (centre, N, S, E and W) which could be averaged together, each of the four habitat assessments were then averaged together in turn to give a value for each variable per transect (and hence per site). Ground and lowstorey vegetation coverage was estimated using a metre square quadrat placed on the ground. Ground vegetation density was estimated using a vegetation density stick (an approximately one metre length of white pvc pipe with fifty black bands) where the number of visible black bands are counted when held at knee and chest height and averaged together. Canopy cover was calculated using the Canopy App (Version 1.0.3, University of New Hampshire) with an Iphone 11 (Model No: MHDH3B/A), held at chest height, with a sensitivity of 100. Canopy height was calculated by measuring the height of the two tallest trees in the vicinity of each habitat assessment (total of 8 trees per transect), using a clinometer, and averaging them (see appendix figure 1).

Tree density was measured by creating three categories of tree: sapling (Diameter at breast height (DBH) <5cm), small tree (DBH 5-15 cm) and large tree (DBH >15cm) using a DBH classification. In each of the four quadrants of each habitat assessment (NE, SE, SW, NW), the distance to the closest individual of each category from the central point was measured. The average distance to each tree type can then be used to calculate the respective density of each tree type, per unit area, using the point centre quarter method (Cottam and Curtis 1956) (see appendix figure 2).

2.3 Statistical Analysis

All analysis was performed using R 4.3.1 (R Core Team 2023) and packages: ggplot2 (Wickham 2016), ggdendro (de Vries and Ripley 2022), vegan (Oksanen et al. 2022), cluster (Maechler et al. 2022), iNEXT (Chao et al. 2014) and MASS (Venables and Ripley 2002).

Using the iNEXT package, sample curves of species richness were created to judge if sampling effort was sufficient. The average number of individuals recorded at each site was approximately 100, and subsequently doubled for an endpoint of 200 individuals for reliability (Hsieh et al. 2016). Whittaker rank abundance plots were created to view overall trends in diversity across sites.

Three alpha diversity indexes were calculated: Simpson's, Shannon's and Fisher's. Simpson's index gives the probability two random individuals belong to different species and is sensitive to dominant species (Simpson 1949). Shannon's index gives the uncertainty of the species identity of a random individual, with more diverse systems having a higher uncertainty (Shannon 1948, Morris et al. 2014). Fisher's index assumes the abundance of species follows a log series distribution (Fisher et al. 1943). True species richness (observed plus undetected) was estimated using two non-parametric estimators: Chao1 and Abundance based Coverage Estimator (ACE). Non-parametric estimators are considered more robust for estimating true species richness than parametric approaches due to the lack of assumptions about the distribution of species abundances and are calculated from the frequencies of rare species (Chao and Chiu 2016). Chao1 is a minimum estimator, based on the frequencies of singletons and doubletons in the sample. ACE, is an estimator based on the concept of sample coverage (how well a sample captures the true assemblage diversity) and estimates the true species richness by defining rare species as those with 10 individuals or fewer in the sample (Chao and Chiu 2016).

Community similarity was compared using UPGMA (bottom up) hierarchal clustering with bray-curtis distance and wards linkage to produce a dendrogram. The two most similar sites/clusters are progressively connected based on pairwise similarity, until all clusters are eventually joined together. A non metric multidimensional scaling (NMDS) plot of each site, with environmental variables fitted, was created using bray-curtis distances to understand similarity and any potential directionality across sites.

Using species richness and abundance as dependent variables, two Generalised Linear Models (GLM's) were carried out to examine associations with environmental parameters. Model normality was checked using residual plots and Shapiro-Wilk tests. Species richness was examined using a model with a Poisson error family and "log" link function which was slightly under dispersed (Theta = 0.80). Poisson distribution is suitable for integer based dependent variables such as count data, provided the model is not overdispersed (Theta > 1.5). Abundance was examined using the same Poisson model for species richness, but was overdispersed (Theta = 2.71) and a negative binomial

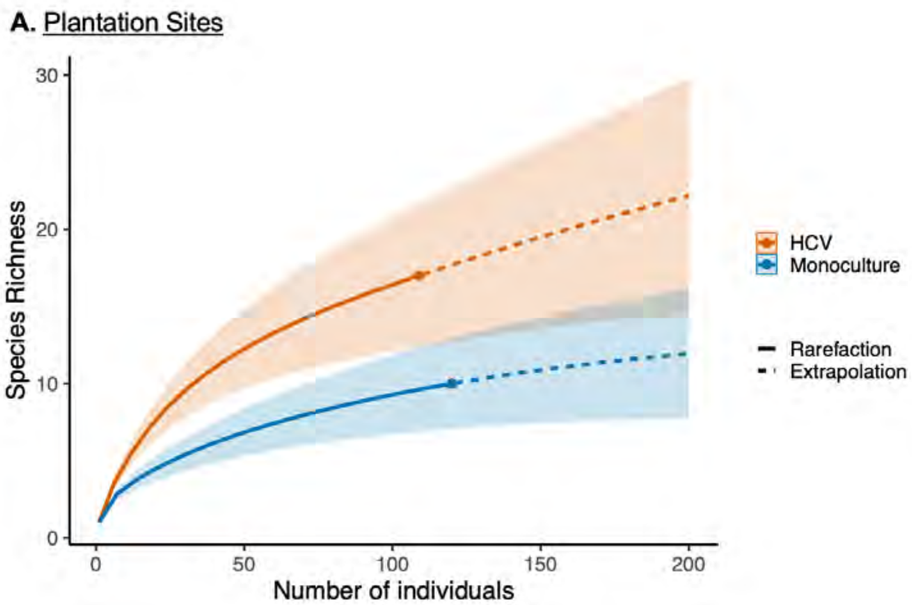
GLM with a "log" link was used instead with a normal dispersion (Theta = 1.23). Independent variables were removed from each model using step wise deletion, based on model normality and Akaike Information Criterion (AIC) values with a change >1 to pick models with the best fit. Environmental parameters with a significant association were subsequently plotted to visualise any trends. For both models ground vegetation depth was chosen to represent vegetation measures because multicollinearity was observed with ground vegetation cover, lowstorey vegetation cover and ground vegetation density. Significance was tested using a standard 95% confidence level.

3.0 Results

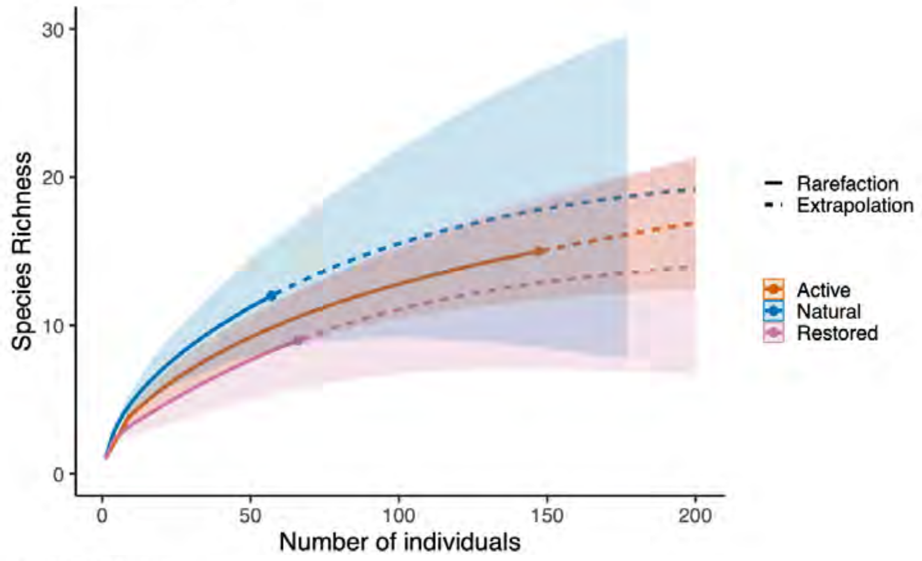
A total of 829 individuals were recorded across 41 species with 14.8 individuals and 4.39 species recorded per sampling event on average. In addition, 65 unknown individuals were recorded, but omitted from final data analysis. Of the 41 species recorded, 16 were singletons and 3 were doubletons. Two species were recorded in all sites, *Appias aegis* (n = 432) and *Parthenos sylvia* (n = 156), which accounted for 52.1% and 18.8% of all individuals recorded respectively.

3.1 Sampling Curves

The gradient of a sampling curve represents the probability of detecting a new species with increasing individuals. The sampling curves for each site (figure 3) show only the restored site at Kaboi lake (figure 4C) appears to reach a plateau, however, the curves for the remaining sites appear to be beginning to plateau.



B. Kaboi Stumping



C. Kaboi Lake

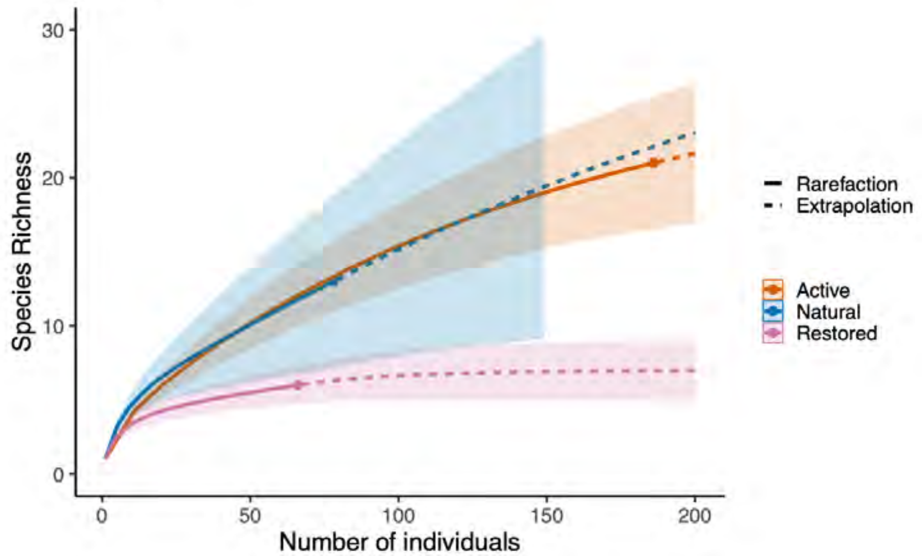


Figure 3. *Sample-based rarefaction/extrapolation (R/E) curves.* Sampling curves are used to show if the sampling effort has been sufficient by visualizing if the curve plateaus/begins to plateau with increasing individuals, indicating the probability of detecting a new species by sampling more individuals is decreasing and therefore whether the sample has captured the majority of the assemblage diversity. The interpolated part of the curve shows the observed species richness, while the extrapolated part shows the predicted species richness up to 200 individuals and 95% confidence limits are included.

3.2 Rank Abundance Plots

Whittaker rank abundance plots are useful for comparing the overall diversity between sites, by ranking each species in descending order of their respective relative abundance (figure 4). A total of 17 and 10 species were recorded in the HCV and monoculture plantation sites respectively, with

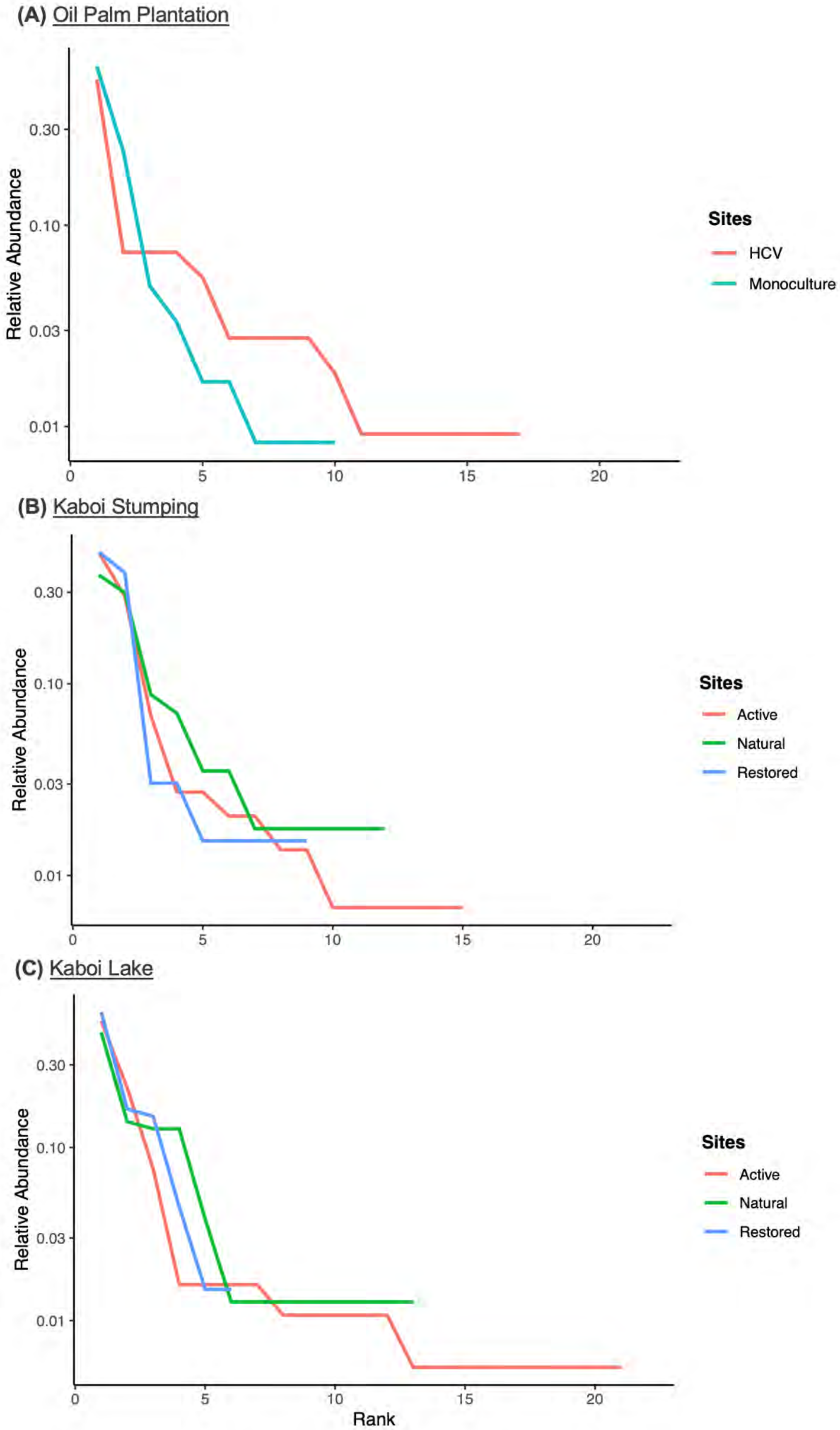


Figure 4. *Whittaker rank abundance plots*. Species recorded at each site are ranked in descending order of their relative abundance. The steepness/shallowness of the curve indicates how even/uneven the distribution of individuals is, and therefore how diverse the community is.

the HCV site showing a shallower and more gradual curve (figure 4A). The active, restored, and natural sites at Kaboi Stumping recorded 15, 12 and 12 species respectively with a fairly even gradient among the respective curves (figure 4B). The active, restored and natural sites at Kaboi Lake recorded 21, 6 and 13 species respectively with a fairly even gradient among the respective curves (figure 4C).

3.3 Alpha Diversity

Alpha diversity is reported in table 1, overall, the active sites had the highest species richness (n=18, SD=4.24) and abundance (n=166.5, SD=27.6) on average while the restored sites had the lowest species richness (n=9, SD=4.24) and abundance (n=66). Of the control sites, species

Table 1. *Alpha Diversity summary information*. Alongside species richness, abundance and rarefied richness (correcting for the minimum number of individuals) three diversity indices were calculated: Simpson's index (inverse), Shannon and Fishers alpha diversity to quantify diversity between sites. Simpsons evenness is also reported, with a higher value indicating fewer dominate species and a more diverse community. True species richness was estimated using two non-parametric indices: a minimum estimator, Chao1, and the abundance based coverage estimator (ACE).

Summary Statistics	Active Restoration		Restored Forest		Natural Forest		Oil Palm Plantation	
	KLA	KSA	KLR	KSR	KLN	KSN	HCV	MNC
Species Summary								
Species Richness (*R*)	21	15	6	12	13	12	17	10
Total Abundance	186	147	66	66	78	57	109	120
Rarefied richness	11.0	9.85	5.73	8.28	10.0	12.0	13.0	7.27
Diversity Indices								
Simpsons Evenness	0.138	0.207	0.396	0.291	0.286	0.345	0.192	0.228
Inverse Simpson Index (1/D)	2.90	3.10	2.38	2.62	3.72	4.14	3.26	2.28
Shannon (*H*)	1.62	1.57	1.16	1.25	1.73	1.79	1.84	1.20
Fisher's Alpha	6.08	4.18	1.60	2.82	4.45	4.64	5.65	2.59
Coverage Estimators								
Chao1	27.0	20.0	7.00	12.3	41.0	17.0	27.5	12.0
Chao1 ± SE	5.38	5.51	2.22	4.09	21.3	5.50	10.5	2.86
ACE	23.8	23.8	9.18	17.8	38.0	21.34	24.6	15.1
ACE ± SE	2.50	3.01	1.72	1.33	3.50	2.24	2.37	1.92

richness increased between the monoculture plantation (n=10), natural sites (n=12.5, SD=0.71) and HCV (n=17); abundance at the monoculture plantation (n=120) was slightly higher than the HCV (n=109) and lowest at the natural sites (n=67.5, SD=14.8).

Simpson's (inverse), Shannon's and Fisher's indices show the monoculture plantation and restored forest sites contained the lowest diversity values (table 1). Values for Simpson's and Shannon's index show the active restoration sites have a lower diversity than the HCV or natural forest sites. Fisher's index values show the HCV and active restoration sites contain a higher diversity than the natural forest.

Chao1 and ACE estimators of species richness (table 1) estimated the restored forest and monoculture plantation contained the lowest true species richness, the active restoration sites and HCV have similar levels of estimated species richness and the highest estimated species richness is in the natural forest site at Kaboi Lake (see table 1).

3.4 Community similarity analysis

A dendrogram analysis showed a dichotomy between the natural forest/KS restored forest cluster and the active/restored/plantation cluster (figure 5). Both active restoration sites formed a cluster which the monoculture plantation shared the highest similarity with. The KL restored forest and

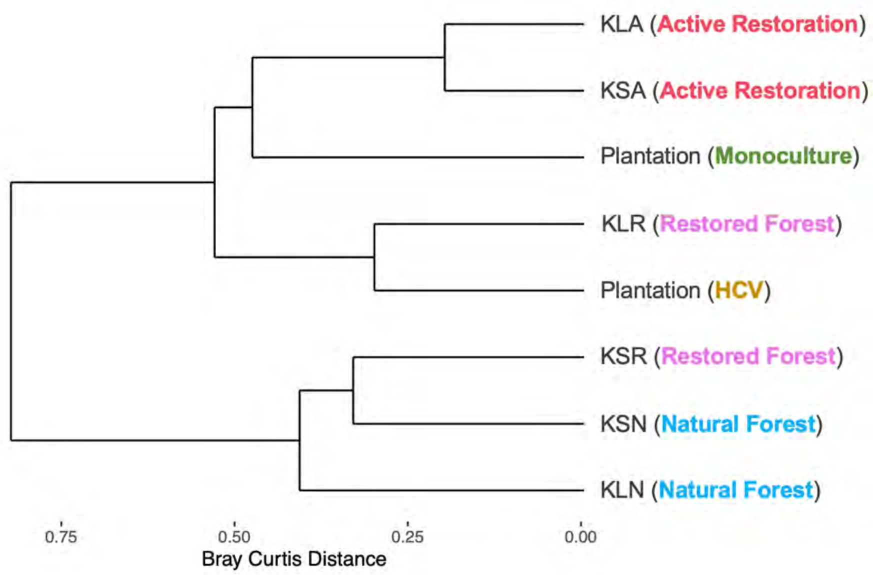


Figure 5. *Hierarchal clustering dendrogram*. Bray-Curtis was used to calculate pairwise distances between sites based on species composition similarity. Clustering was carried out using wards linkage method (UPGMA) where the most similar sites are progressively joined together in clusters, until all clusters are joined together.

HCV site formed a cluster, sharing the highest levels of similarity.

An NMDS plot ordinating community data further shows similarity between sites (figure 6). Overall, the study site types are shown to be associated amongst each other, with the strongest similarity appearing between both active forest restoration sites. Along the NMDS1 axis, there is a clear distinction between forest and plantation areas, with the monoculture plantation site appearing the most distinct. Natural and restored forest areas appear loosely associated and the HCV plantation site and active forest restoration sites appear somewhere inbetween. Environmental variables were fitted, with temperature increasing directionally from the forest towards the plantation. Canopy cover and canopy height strongly directionally increase towards the natural forest, while large tree density increases directionally towards the natural forest and plantation median. Small tree density and sapling density increase directionally towards the restored and natural forest median.

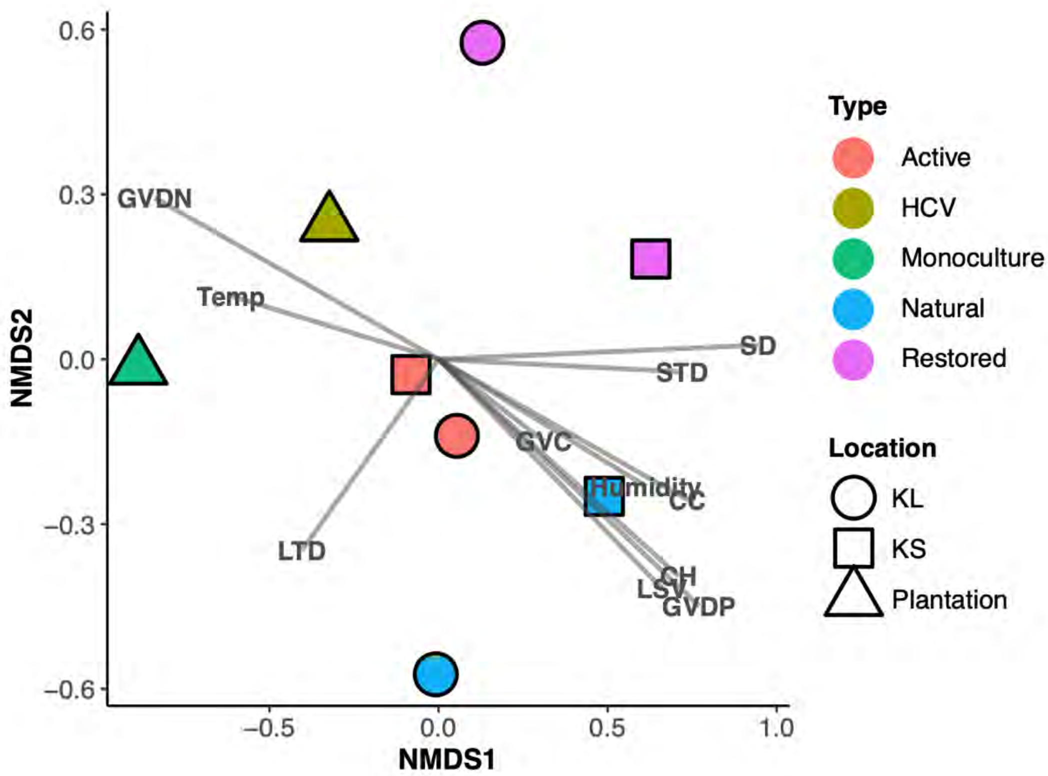


Figure 6. *NMDS ordination.* Pairwise distances were calculated using bray curtis. The more similar two sites are in butterfly species composition the closer together they appear. Environmental variables were fitted (values increase in respective directions across the plot): ground vegetation cover (GVC), lowstorey vegetation cover (LSV), ground vegetation density (GVDN), ground vegetation depth (GVDP), temperature (temp), humidity, canopy cover (CC), canopy height (CH) and the densities of saplings (SD), small trees (STD) and large trees (LTD) respectively.

3.5 Environmental variable modelling

A Poisson GLM using species richness as the dependent variable and environmental characteristics as the independent (explanatory) variables showed species richness was significantly associated with temperature and sapling density (table 2 model 1). Included in the model, but not significantly associated, were: canopy height, canopy cover and large tree, small tree and sapling density. A negative binomial GLM using abundance as the dependent variable and environmental characteristics as the explanatory variables showed abundance was significantly associated with: temperature, canopy cover and large tree, small tree and sapling density (table 2 model 2).

Table 2. *Environmental parameter GLM results*. Two GLM models were carried out to test the association between species richness (model 1) and abundance (model 2) with environmental variables. The association of species richness with environmental variables was examined using a Poisson GLM with a slight under dispersion (theta = 0.796) and a pseudo R^2 of 0.447. The association of abundance with environmental variables was examined using a negative binomial GLM with a normal dispersion (theta = 1.23) and a pseudo R^2 of 0.541. The estimate indicates the expected change in the y value (dependent variable) given an increase of 1 in the x value (independent variables). A significant p value is denoted by *. The residual deviance degrees of freedom for both models was 32.

Dependent Variable	Parameter	Estimate	SE (±)	z value	p value
(Model 1) Species Richness	*Intercept*	*-3.623*	*2.218*	*-1.634*	*0.102*
	Temperature	0.128	0.060	2.152	0.031*
	Canopy Height	0.074	0.074	1.001	0.317
	Large Tree Density	-13.870	7.547	-1.838	0.066
	Small Tree Density	5.332	3.063	1.741	0.082
	Sapling Density	-6.252	3.024	-2.068	0.038*
	Canopy Cover	0.016	0.009	1.746	0.081
(Model 2) Abundance	*Intercept*	*-5.384*	*2.150*	*-2.504*	*0.012**
	Temperature	0.226	0.057	3.972	<0.001*
	Large Tree Density	-11.554	4.259	-2.713	0.007*
	Small Tree Density	4.454	2.093	2.128	0.033*
	Sapling Density	-5.790	1.665	-3.477	0.001*
	Canopy Cover	0.031	0.009	3.484	<0.001*

The relationship of significant associations between species richness and abundance with environmental variables from the GLM output (table 2) were plotted (figure 7). Species richness (figure 7A) and abundance (figure 7B) with temperature show a positive correlation, however, the

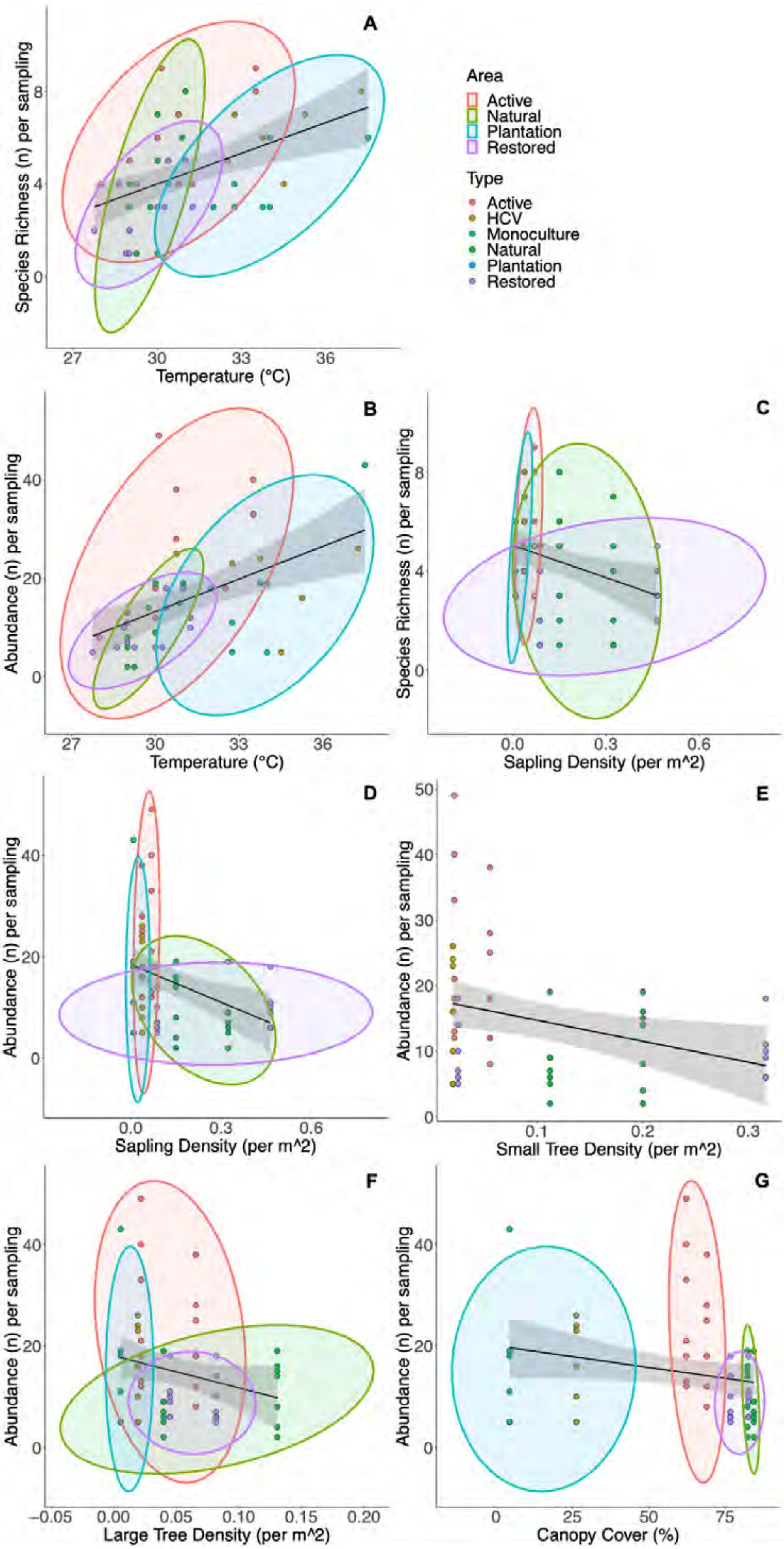

A
B
C
D
E
F
G
Species Richness (n) per sampling
Abundance (n) per sampling
Temperature (°C)
Sapling Density (per m^2)
Small Tree Density (per m^2)
Large Tree Density (per m^2)
Canopy Cover (%)
Area
Active
Natural
Plantation
Restored
Type
Active
HCV
Monoculture
Natural
Plantation
Restored

Figure 7. *Scatter plots of significant GLM environmental variables.* Associations of environmental parameters with species richness and abundance were examined with a Poisson GLM and a Negative Binomial GLM respectively (see table 2). Significant environmental variables from the models were plotted with a fitted regression line (with 95% confidence intervals) to illustrate any trends. Species richness and abundance were both significantly associated with temperature (**A** and **B**) and sapling density (**C** and **D**). Abundance was also significantly associated with: small tree density (**E**), large tree density (**F**) and canopy cover (**G**). Ellipses were fitted where feasible to better illustrate the overall trends regarding different study areas.

GLM estimate output was near zero (table 2) for both models. Natural and restored forest are associated with relatively cooler temperatures while active restoration and plantation sites are associated with relatively warmer temperatures. Regarding sapling density, species richness (figure 7C) and abundance (figure 7D) show a negative association which correlates with the negative estimate output for both models (table 2). Natural and restored forest are associated with a relatively high sapling density while plantation and active restoration sites are associated with a relatively low sapling density. Small tree density had a positive estimate output (table 2) with abundance, but figure 7E shows a negative association. Small tree and large tree density had positive and negative estimate outputs respectively (table 2) regarding abundance, however, both show a slight negative trend (figure 7E and 7F). Active restoration and plantation sites are associated with relatively low small and large tree densities, while restored and natural forests appear associated with relatively higher densities. Canopy cover was a significant variable regarding abundance (table 2) but the estimate output was near zero, when plotted (figure 8G) there appears to be no clear trend with a relatively horizontal curve.

4.0 Discussion

This study recorded approximately 45% of the known butterfly species in the lower Kinabatangan; compared to a previous study (Owen, unpublished 2019) who used a combination of transect walks and bait traps to quantify species richness in the natural forest, over a six month period, and recorded 92 species. Another study (Webb, unpublished 2020) compared species diversity between plantation, natural forest and two restored forest sites along the Kinabatangan river (including Kaboi Stumping restored) over a period of five weeks, recording 33 species of which this study shared 15.

Sampling curves suggested most species richness had been captured, however, the study by Owen (unpublished, 2019) provides the most comprehensive assessment of butterfly species richness in the same locality; which this study recorded half as many species. At the restored KS site, Webb (unpublished, 2020) identified 18 species whereas this study only recorded 12 despite sampling effort being approximately the same. Given the lack of a clear plateau in the sampling curves and comparisons of observed species richness with previous studies in the local area, its likely this study under sampled assemblages. Further sampling effort would have likely yielded more species and this study only took place for seven out of the planned ten weeks, due to unforeseen animal and weather events.

4.1 Diversity and community similarity between restoration and control sites

The active sites contained the highest recorded species richness and abundance on average, while the restored sites contained the lowest, with the natural forest and plantations sites intermediary. However, the restored forest site at KS recorded the same number of species as the natural forest sites and the average was brought down by the restored forest site at KL which recorded half as many species. This contrasts with the results of Webb (unpublished, 2020) who recorded double the number of species in the restored KS site compared to the natural forest, but similarly to this study did find species richness was higher in the restored forest than the monoculture plantation.

The rank abundance curves suggested the active restoration sites and restored forest sites were less diverse and contained more dominant species, which loosely correlated with measures of evenness. Alpha diversity values suggested the monoculture and restoration sites contained the lowest diversity; while the active restoration, natural forest and HCV plantation sites had similarly high diversity. The evenness of the active restoration and plantation sites was lower than the restored and natural sites, which coincides with results from Kwatrina et al. (2018) which found compared to forests, plantations had lower evenness and contained more dominant species.

Community composition comparisons showed there was a clear distinction between plantation and forest sites, with the active restoration sites inbetween. This shows that the forest and plantations are made of different communities, with evenness and diversity metrics suggesting the forests contain fewer dominant individuals. The restored forests and monoculture were comparatively low in diversity but contained different communities illustrating their different biodiversity values. Cluster analysis suggested the restored forest at KS, shared a high level of similarity with the natural forest sites suggesting community composition and therefore perhaps the site itself has been restored to a relatively natural state. Whereas the restored site at KL, shared more similarity with the active and plantation sites; sharing the closest similarity with the HCV plantation forest reserve. One reason could be because of the higher level of susceptibility to flooding for both sites and the presence of similar flood tolerant plants and therefore butterfly compositions.

4.2 Forest restoration directionality trends

Overall, the measures of species richness, abundance and diversity indices didn't show a clear overall increase with directionality between the active, restored and natural forest sites nor with age. The restored forest at KL was four years older than the restored forest at KS, yet contained half the species richness and was lower in all diversity measures and richness estimators. This could be due to the differences in the type of site, KL restored being a freshwater swamp which is seasonally flooded up to approximately three metres; compared to KS restored which is a riparian forest experiencing more of the year relatively dry. The restored forest at KS did contain the same species richness as the natural forest at KS and a greater abundance of individuals, but all diversity metrics were lower. Community ordination analysis didn't reveal a directionality either in species composition changes between active, restored and natural forest. The NMDS plot in this study didn't reveal a directionality of progression between the active, restored and natural forest; but did show the natural forest and plantation areas are clearly distinct with the restored and particularly the active sites somewhere intermediary in species composition.

Overall, these results contrasts with other tropical restoration studies, which typically find species richness and diversity increases as well as directional community compositional changes with restoration age up to the natural forest (Itioka et al. 2015, Korkiatupa et al. 2023). For example, the NMDS plot in Itioka et al. (2015) shows a directionality in community composition between early stages of succession to restored forests to natural forest which this study did not show. However, these studies benefited both from a greater number of available restored forest plots and varying ages which could provide a greater insight on butterfly assemblage changes this study was not afforded. Itioka et al. (2015) studies succession plots from three to sixty years of restoration age and Korkiatupa et al. (2023) involved a multi decade study, studying both new and old restoration plots dating to thirty years old. These studies were also able to compare restoration plots to primary forest, which this study was unable to due to the intensity logging in the lower Kinabatangan, leaving the true identity of the natural forest community composition uncertain.

4.3 Environmental variable associations with species richness and abundance

Modelling with environmental parameters suggested species richness and abundance would increase with temperature and decrease with increasing tree densities. Canopy cover was suggested to be an explanatory variable for abundance, but the nature of the association was uncertain. Other studies have shown species richness increases with canopy cover (Rija 2022), however, canopy cover was not an explanatory variable for species richness in this study and the trend was slightly negative regarding abundance if anything at all. Species richness and abundance were recorded at higher temperatures and lower tree densities, which are characteristics associated with the HCV plantation and active restoration sites which would have experienced greater light intensities from the reduced canopy cover.

Light is an important abiotic factor in tropical forest dynamics determined by canopy cover (Hill et al. 2001). Tropical forest canopy is a mosaic of regenerating gaps from tree fall creating heterogenous areas within the forest landscape with varying microclimate effects such as light intensity (Hamer et al. 2003). The increased habitat heterogeneity can provide new niches for species to colonise and potentially lead to an increased species richness (Cleary and Genner 2006). Moreover, selectively logged forests are typically more homogenous than primary forests, due to the more extensive canopy gaps (from large tree removal) leading to rapid colonisation by a few pioneer species (Hamer et al. 2003). This could explain why a higher species richness and abundance were recorded at the active restored sites, because the site is surrounded by forest and the more heterogenous structure could create more niches for species to exploit.

There is a clear difference between community composition of the active restored sites and the natural forests, suggesting this increase in species richness might not necessarily be beneficial. A result of canopy loss is the loss of understorey forest dependent species with more restrictive habitat requirements and of greater conservation value (Hill 1999) and the increase in widespread generalist species (Hill et al. 2001). Furthermore, the NMDS plot with environmental variables fitted, showed community compositional changes along gradients of climate, canopy and tree density variables further demonstrating the sensitives of different species to different microclimates (Hamer et al. 2003).

4.4 Experimental design critique, implications and future directions

The sampling method used was a modified walk and count pollard method (Pollard 1977), which has been suggested as a useful tool for monitoring tropical butterfly diversity (Basset et al. 2011, Sparrow et al. 1994) Having been developed for temperate areas with relatively low diversity, the method is criticised in the tropics because of: large numbers of species, high levels of species similarity, lower light intensities below the canopy and dense vegetation impeding spotting ability (Walpole and Sheldon 1999). Pollard walks also assume a constant level of detectability for each butterfly which due to individual behaviour (e.g. conspicuousness) or habitat features (e.g. high canopy cover) is rarely the case in the tropics (Pellet et al. 2012). Transects count can also be affected by the openness of the habitat (Thomas 1983) and the higher species richness and abundance observed at the active and plantation sites may just have been the result of the more open landscape making spotting easier. As the study period progressed, “on the wing” identification was increasingly relied on as identification skill increased (Sparrow et al. 1994), but the use of genera for sampling/monitoring is easier and more reliable (Cleary 2004) and would be suitable for a long-term monitoring program involving different participants every year.

Light levels below and above the canopy can lead to different butterfly assemblages (Hill et al. 2001) and ground based approaches (such as this study) underrepresent canopy specialists (Walpole and Sheldon 1999). Bait trapping can sample both ground and canopy assemblages and should be used in tandem in the tropics with pollard walks (Sparrow et al. 1994). Owen (unpublished, 2019) identified ten species not observed through transect walks but with bait trapping only, highlighting species diversity this study isn't capturing. However, Owen (unpublished, 2019) also found long tailed macaques (*Macaca fascicularis*) heavily interfered with trapping efforts, and future studies should attempt to see if bait trapping could be a feasible addition to restoration monitoring.

Butterflies have specific plant preferences for hosting larvae and feeding and butterfly diversity is suggested to correlate with plant and flower diversity (Kremen et al. 1992). A lower observed butterfly diversity in restored forest areas could indicate a lower diversity of host plants and recovery stagnation, which would be understandable given the relative homogeneity from using a few hardy, rapid pioneer species to restore the canopy of sites. Its unclear from this study, but if butterfly diversity and community composition doesn't improve in restored sites, active planting of climax species to increase plant diversity might be necessary (Chazdon et al. 2009). Moreover, an assessment of plant species would reveal if these climax species are indeed growing, and the lack of apparent diversity and compositional changes is from their naturally slow growth.

This study also compared individual sites under broad groupings such as “restored forest” or “active site” which was necessary given the limited sample size; yet the analysis assumes a similarity of the study areas (Korkiatupa et al. 2023) and grouping together different habitats such as freshwater swamp and riparian forest might not be very suitable. However, studies like Korkiatupa et al. (2023) studied dozens of restoration sites, whereas this study could only look at four and so while not necessarily suitable the groupings were necessary. In addition, this study attempted to use GLM's to identify any environmental predictors of species richness and abundance but didn't identify any clear biologically relevant associations; while less powerful, canonical correspondence analysis can be useful for explaining community composition variance with environmental variables and predicting restoration effects (Nyafwono et al. 2015) and could have been attempted instead.

Finally, studies typically standardise by sampling effort, sample size (rarefaction) or sample coverage (how well your sample captures true community diversity including undetected species) (Roswell et al. 2021). This study used the simplest approach, which is to standardise by sampling effort, however, this method is often criticised for underestimating community diversity; because smaller samples will contain fewer individuals and therefore fewer species by chance (Roswell et al. 2021). Increasingly consensus has suggested samples should be standardised by coverage and it would be useful for a future study to compare the diversity results from standardising by different methods, to determine the most accurate way to quantify and compare butterfly diversity.

4.5 Conclusion

Overall, the results of this study didn't suggest there was a clear effect of forest restoration on measures of alpha diversity and directionality of butterfly diversity and community composition; nor did this study identify any biologically clear associations between species richness and abundance with environmental predictors of forest restoration. However, this study provides the first butterfly data set on forest restoration under the new Regrow Borneo project and provides the foundational data for a longer-term monitoring program which should expand to cover the new restoration sites and expand on the methodology used in this study. Forest restoration is an important strategy for mitigating climate change, preserving biodiversity and long term assessments of restoration success are important for informing future efforts.

Acknowledgements

Thankyou to Amaziasizamoria Jumail and Dr Pablo Orozco ter Wengel for suggesting this project idea to me and providing guidance and support. Many thanks to Luke Davies and Tyler Cuddy for help with statistical analysis and the use of R. A massive gratitude is shared to the local Field Assistants: Alut, Mit, Lee, Zul, Pian and Otto without whom the fieldwork for this project would not have been possible. Thank you to all the students at DGFC who helped me with my sampling, particularly Hannah Shapland who was a big help throughout this project and helped carry out some very fun and memorable habitat assessments. Finally, a big thankyou to my supervisor Benoit Goossens and the Regrow Borneo project for affording me the opportunity to carry out this study.

References

- Abram, N. K. et al. 2014. Synergies for improving oil palm production and forest conservation in floodplain landscapes. *PLOS ONE* 9(8): e106391. DOI: https://doi.org/10.1371/journal.pone.0095388
- Achard, F., Eva, H. D., Stibig, H. J., Mayaux, P., Gallego, J., Richards, T. and Malingreau, J. P. 2002. Determination of Deforestation rates of the worlds humid tropical forests. *Science* 297(5583), pp. 999-1002. DOI: 10.1126/science.1070656

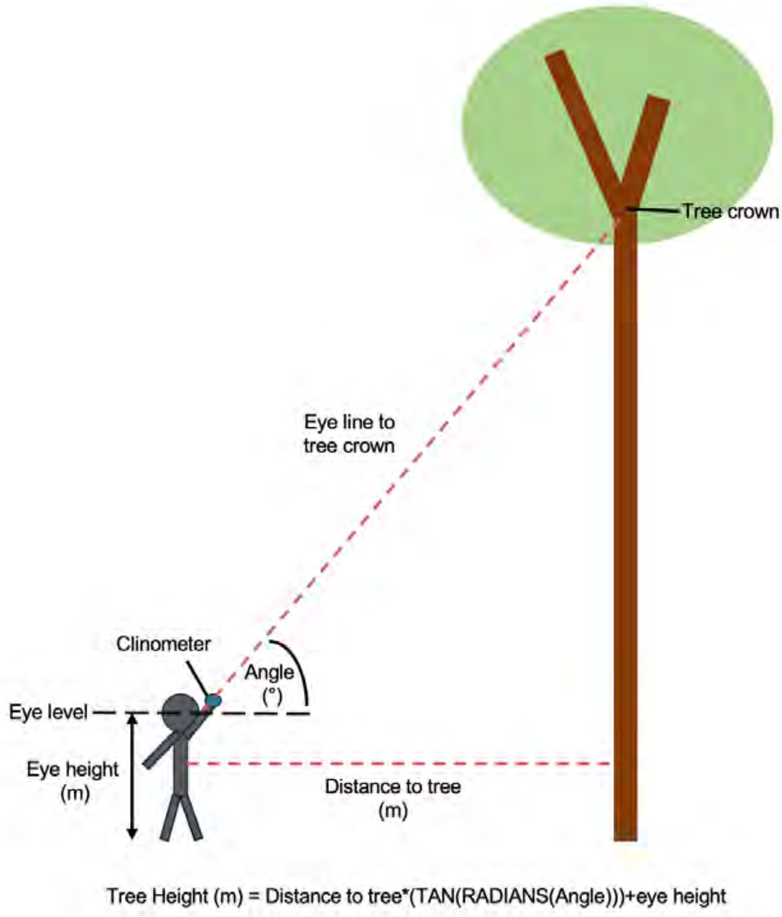
- Ancrenaz, M., Goossens, B., Gimenez, O., Sawang, A. and Lackman-Ancrenaz, I. 2004. Determination of ape distribution and population size using ground and aerial surveys: a case study with orang-utans in Lower Kinabatangan, Sabah, Malaysia. *Animal Conservation* 7(4), pp. 375-385. DOI: https://doi.org/10.1017/S136794300400157X
- Basset, Y. et al. 2011. Comparison of rainforest butterfly assemblages across three biogeographical regions using standardised protocols. *The Journal of Research on the Lepidotera* 44: 17-28.
- Benedick, S., Hill, J. K., Mustaffa, N., Chey, V. K., Maryati, M., Searle, J. B., Schilthuizen, M. and Hamer, K. C. 2006. Impacts of rain forest fragmentation on butterflies in northern Borneo: species richness, turnover and the value of small fragments. *Journal of Applied Ecology* 43(5), pp. 967-977. DOI: https://doi.org/10.1111/j.1365-2664.2006.01209.x
- Bierregaard, R. O., Lovejoy, T. E., Kapos, V., Santos, A. A. and Hutchings, R. W. 1992. The Biological dynamic of tropical rainforest fragments. *Bioscience* 42(11), pp. 859-866. DOI: https://doi.org/10.2307/1312085
- Bonebrake, T. C., Ponisio, L. C., Boggs, C. L. and Ehrlich, P. R. 2010. More than just indicators: A review of tropical butterfly ecology and conservation. *Biological Conservation* 143: 1831-1841. DOI: https://doi.org/10.1016/j.biocon.2010.04.044
- Brookfield, H. and Byron, Y. 1990. Deforestation and timber extraction in Borneo and the Malay Peninsula. *Global Environmental Change* 1(1), pp. 42-56. DOI: https://doi.org/10.1016/0959-3780(90)90006-U
- Chao, A. and Chiu, C. H. 2016. Species Richness: Estimation and Comparison. *Wiley StatsRef: Statistics Reference Online*. DOI: https://doi.org/10.1002/9781118445112.stat03432.pub2
- Chazdon, R. L. et al. 2009. The potential for species conservation in tropical secondary forests. *Conservation Biology* 23(6), pp. 1406-1417. DOI: 10.1111/j.1523-1739.2009.01338.x
- Cleary, D. F. R. 2004. Assessing the use of butterflies as indicators of logging in Borneo at three taxonomic levels. *Journal of Economic Entemology* 97(2), pp. 429-435. DOI: 10.1093/jee/97.2.429
- Cleary, D. F. R. and Genner, M. J. 2006. Diversity patterns of Bornean butterfly assemblages. *Biodiversity and Conservation* 15, pp. 517-538. DOI: https://doi.org/10.1007/s10531-005-2353-4
- Cottam, G. and Curtis, J. T. 1956. The use of Distance Measures In Phytosociological Sampling. *Ecology* 37(3), pp. 451-460. DOI: https://doi.org/10.2307/1930167
- de Vries, A. and Ripley, B. D. 2022. ggdendro: Create Dendrograms and Tree Diagrams Using ‘ggplot2’.

- Chao, A., Gotelli, N. J., Hiseh, T. C., Sander, E. L., Ma, K, H., Colwell, R. K. and Ellison, A. M. 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecological Monographs. 84:45-67.
- Fisher, B., Edwards, D. P., Larsen, T. H., Ansell, F. A., Hsu, W. W., Roberts, C. S., Wilcove, D. S. 2011. Cost effective conservation calculating biodiversity and logging trade offs in southeast Asia. *Conservation letters* 4(6), pp. 443-450. DOI: https://doi.org/10.1111/j.1755-263X.2011.00198.x
- Fisher, R. A., Corbet, A. S. and Williams, C. B. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology* 12(1), pp. 42-58. DOI: https://doi.org/10.2307/1411
- Gaveau, D. L. A., Sloan, S., Molidena, E., Husna, Y., Sheil, D., Abram, N. K., ... Meijaard, E. 2014. Four decades of forest persistence, clearance and logging on Borneo. *PLOS ONE* 9(7):1-11. DOI: https://doi.org/10.1371/journal.pone.0101654
- Hamer, K. C., Hill, J. K., Benedick, S., Mustaffa, N., Sherratt, T. N., Maryati, M. and Chey, V. K. 2003. Ecology of butterflies in natural and selectively logged forests of northern Borneo: the importance of habitat heterogeneity. *Journal of Applied Ecology* 40(1):150-162. DOI: https://doi.org/10.1046/j.1365-2664.2003.00783.x
- Hill, J. K. 1999. Butterfly spatial distribution and habitat requirements in a tropical forest: impacts of selective logging. *Jounral of Applied Ecology* 36(4), pp. 564-572. DOI: https://doi.org/10.1046/j.1365-2664.1999.00424.x
- Hill, J. K., Hamer, K. C., Tangah, J. and Dawood, M. 2001. Ecology of tropical butterflies in rainforest gaps. Oecologia. 128: 294-302. DOI 10.1007/S004420100651
- Itioka, T. et al. 2015. Chronosequential changes in species richness of forest- edgedwelling butterflies during forest restoration after swidden cultivation in a humid tropical rainforest region in Borneo. *Journal of Forest Research* 20(1):125-134. DOI: https://doi.org/10.1007/s10310-014-0444-3
- Korkiatupa, E. et al. 2023. Recovery patterns in community composition of fruit feeding butterflies following 26 years of active forest restoration. *Ecosphere* 14(5), e4514. DOI: https://doi.org/10.1002/ecs2.4514
- Kremen, C. 1992. Assessing the indicator properties of species assemblages for natural areas monitoring. *Ecological applications* 2(2): 203-217. DOI: 10.2307/1941776
- Kwatrina, R. T., Santosa, Y., Bismark, M. and Santoso, N. 2018. Ecological impacts of oil palm plantation on butterfly and bird species diversity. *Jurnal Manajemen Hutan Tropika* 24(1), pp. 23-31. DOI: 10.7226/jtfm.24.1.23
- Lamb, D. Erskine, P. D. and Parrotta, J. A. 2005. Restoration of degraded tropical forest landscapes. *Science* 310(5754), pp. 1628-1632. DOI: 10.1126/science.1111773
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. and Hornik, K. 2022. Cluster: Cluster Analysis Basics and Extensions.

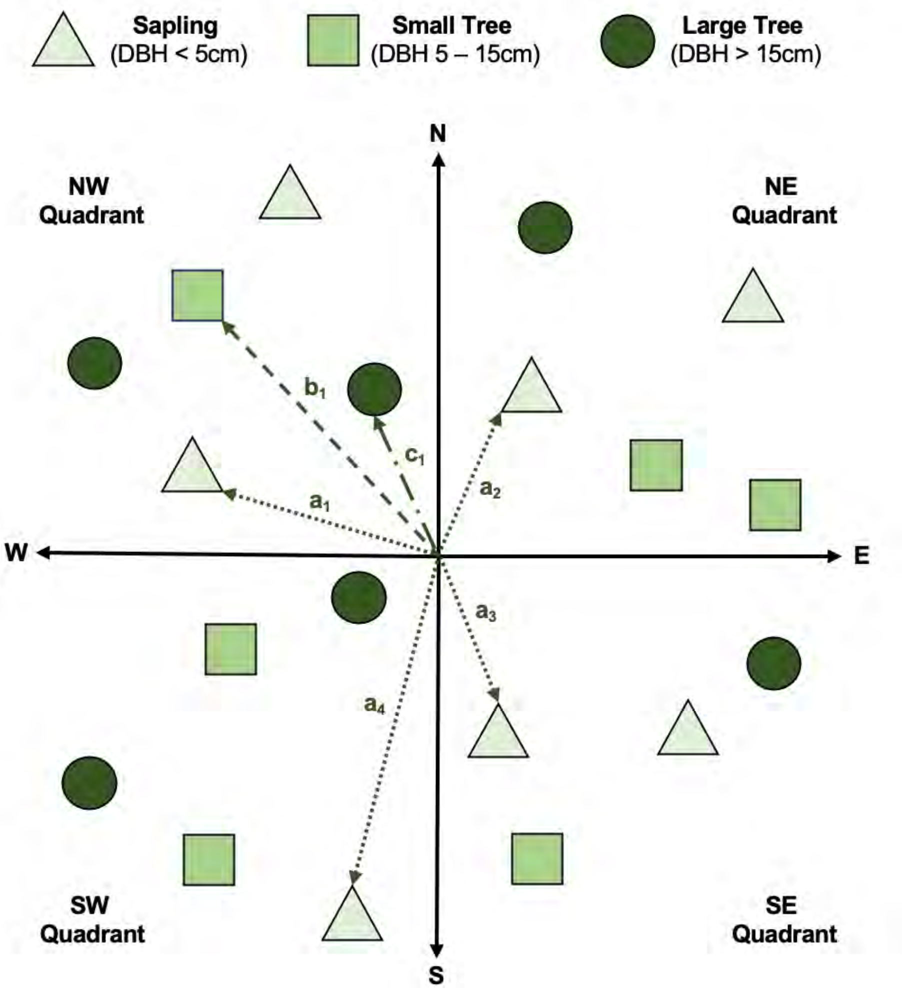
- Morris, K. E. et al. 2014. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Explanatories. *Ecology and Evolution* 4(18), pp. 3514-3524. DOI: https://doi.org/10.1002/ece3.1155
- Nyafwono, M., Valtonen, A., Nyeko, P., Owiny, A. A. and Roininen, H. 2015. Tree community composition and vegetation structure predict butterfly community recovery in a restored afrotropical rainforest. *Biodiversity and Conservation*. 24, pp. 1473-1485. DOI: https://doi.org/10.1007/s10531-015-0870-3
- Otsuka, K. 2001. *A Field Guide to the Butterflies of Borneo and South East Asia*. Kota Kinabalu: Hornbill Books.
- Oksanen, J. et al. 2022. vegan: Community Ecology Package.
- Pellet, J., Bried, J. T., Parietti, D., Gander, A., Heer, P. O., Cherix, D. and Arlettaz, R. 2012. Monitoring butterfly abundance beyond pollard walks. *PLOS ONE* 7(7), e41396. DOI: https://doi.org/10.1371/journal.pone.0041396
- Pollard, E. 1977. A method for assessing changes in the abundance of butterflies. *Biological Conservation* 12(2):115-134. DOI: https://doi.org/10.1016/0006-3207(77)90065-9
- Owen, J. 2019. Biodiversity of butterflies in the Lower Kinabatangan Wildlife Sanctuary and the influence of habitat variables. Available from: https://www.dgfc.life/home/pty-reports/ [Accessed: 15/07/2023]
- R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org
- Regrow Borneo. 2023. Regrow Borneo Our Approach. Available at: https://regrowborneo.org/our-approach-2 [Accessed: 20/08/2023]
- Rija, A. A. 2022. Local habitat characteristics determine butterfly diversity and community structure in a threatened Kihansi gorge forest, Southern Udzungwa Mountains, Tanzania. *Ecological Processes* 11, 13. DOI: https://doi.org/10.1186/s13717-022-00359-z
- Roswell, M. Dushoff, J. and Winfree, R. 2021. A conceptual guide for measuring species diversity. *Oikos* 130(3), pp. 321-338. DOI: https://doi.org/10.1111/oik.07202
- Scriven, S. A., Beale, C. M., Benedick, S. and Hill, J. K. 2016. Barriers to dispersal of rain forest butterflies in tropical agricultural landscapes. *Biotropica* 49(2), pp. 206-216. DOI: https://doi.org/10.1111/btp.12397
- Shannon, C. E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27, pp. 379-423, 623-656.
- Simpson, E. H. 1949. Measurement of Diversity. *Nature* 163, pp. 688. DOI: http://dx.doi.org/10.1038/163688a0
- Sparrow, H. R., Sisk, T. D., Ehrlich, P. R. and Murphy, D. D. 1994. Techniques and guidelines for monitoring neotropical butterflies. *Conservation Biology* 8(3):800-809.

- Stork, N. E., Srivastava, D. S., Watt, A. D. and Larsen, T. B. 2003. Butterfly diversity and silvicultural practice in lowland rainforests of Cameroon. *Biodiversity and Conservation* 12, pp. 387-410
- Thomas, J. A. 1983. A quick method for estimating butterfly numbers during surveys. *Biological Conservation* 27: 195-211.
- Venables, W. N. and Ripley, B., D. 2002. Modern Applied Statistics with S. Fourth Edition. Springer: New York.
- Webb, K. 2020. Butterflies (Order: Lepidoptera) as a biological indicator: Comparing diversity and species composition between three forest sites and an oil palm plantation in the Lower Kinabatangan Wildlife Sanctuary, Sabah, Malaysian Borneo. Available from: https://www.dgfc.life/home/pty-reports/ [Accessed: 20/10/2022]
- Wickham, H. 2016. Elegant Graphics for Data Analysis. Springer-Verlag: New York.

Supplementary Material



Appendix figure 1. *Tree height calculations.* A standard point on each tree should be chosen to measure to, this study used the first major tree crown (branching). The recorded angle of the clinometer, distance of the surveyor from the tree and the surveyors eye height can be used to calculate tree height using the provided formula in Microsoft Excel.



$$\textbf{Equation 1}: \text{ Average distance}(xAv) = \frac{x_1 + x_2 + x_3 + x_4}{4}$$

$$\textbf{Equation 2}: \text{ Density} = \frac{1}{(xAv)2}$$

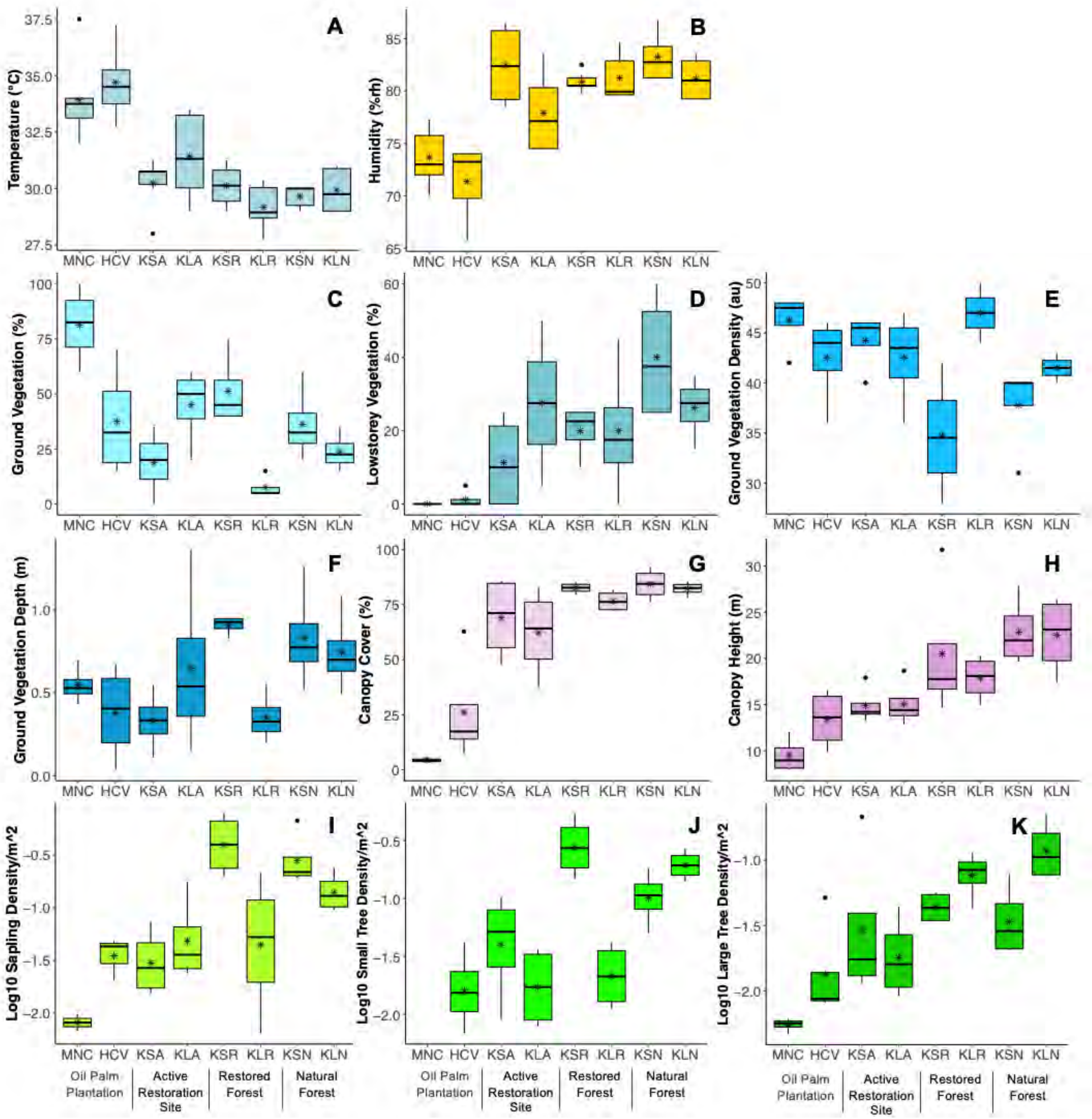
Appendix Figure 2. *Point Centre Quarter Method.* Three categories of tree were created based on diameter at breast height (DBH) sizes: Sapling (DBH < 5cm), Small Tree (DBH 5 – 15cm) and Large Tree (DBH > 15cm). Each habitat assessment had four quadrants: North East, North West, South East and South West. In each quadrant, the distance (x) to the nearest tree of each category to the central point is measured; these distances are then averaged to calculate a mean tree distance (x_{Av}) (equation 1) which are used to calculate a density of trees per unit area (equation 2). Depending on which density category calculation is desired, x could be a, b or c distances.

Appendix table 1. *Raw count data*. Count data for each species recorded at each site across a total of seven weeks of sampling.

Species	HCV	MNC	KSN	KSR	KSA	KLA	KLR	KLN
Appias aegis	58	74	21	32	71	100	40	36
Appias lyncida	0	0	0	0	0	1	0	0
Athyma nefte	2	0	0	1	0	2	1	1
Cepora iudith	1	0	0	0	0	0	0	0
Cirrochroa satellita	0	0	4	0	0	0	0	0
Cupha erymanthis	1	0	0	0	0	0	0	0
Cuphia arias	8	0	2	0	10	14	10	1
Cyrestis theresae	0	1	0	0	0	0	0	0
Doleschallia bisaltide	1	0	0	0	0	0	0	0
Drupadia ravindra	0	0	0	2	0	0	0	10
Eupolea mulciber	0	0	0	0	0	1	0	0
Eurema sp	1	6	0	0	4	3	0	0
Euthalia/Tanaecia sp	0	0	1	1	0	1	0	0
Graphium agamemnon	0	2	0	0	0	1	0	0
Graphium sarpedon	0	0	0	0	2	0	0	0
Hebomoia glaucippe	0	0	1	0	0	0	0	0
Hypolimnas bolina	0	0	0	0	0	2	0	0
Hypolimnas misippus	6	4	0	0	1	3	1	1
Idea stoli	0	0	5	0	1	1	0	10
Ideopsis vulgaris	3	0	1	0	3	0	3	0
Jamides celeno	0	0	0	0	0	1	0	0
Junonia hedonia	3	1	0	0	0	0	0	0
Junonia iphita	8	28	0	0	0	1	0	0
Koruthaialos sindu	0	0	1	0	0	0	0	0
Leptosia nina	0	1	0	0	0	0	0	1
Moduza procris	1	0	0	0	0	0	0	0
Mycalesis anapita	0	0	0	0	2	3	0	1
Nacaduba berenice	0	0	0	1	0	0	0	0
Neptis duryodana	3	2	0	0	3	2	0	0
Panidta sinope	0	0	1	0	1	2	0	1
Papilio nephelus	1	0	1	2	1	3	0	0
Paralaxita telesia	0	0	0	0	0	0	0	3
Parantica crowyeli	0	0	0	0	0	0	0	1
Parthenos sylvia	8	1	17	25	42	41	11	11
Polyura athamas	0	0	0	0	0	1	0	0
Thaumantis klugius	0	0	0	0	1	0	0	0
Troides amphrysus	1	0	2	1	0	2	0	0
Vindula erota	3	0	0	0	0	1	0	0
Xanthotaenia busiris	0	0	0	0	0	0	0	1
Ypthima fasciata	0	0	0	1	4	0	0	0
Zeuxidia doubledayi	0	0	0	0	1	0	0	0

Appendix table 2. *Habitat data*. Summary habitat data recorded for each environmental parameter at each site with standard error. Small tree density was not recorded at Pendirosa plantation (MNC). Four habitat assessments were conducted per 200m of transect which were averaged together and rounded where appropriate.

Environmental Variables (± SE)	**Oil Palm Plantation**		**Active Restoration Site**		**Restored Forest**		**Natural Forest**	
	MNC	HCV	KSA	KLA	KSR	KLR	KSN	KLN
Ground Vegetation								
Ground Vegetation (%)	80 (±9)	40 (±13)	20 (±7)	45 (±9)	50 (±8)	10 (±3)	35 (±9)	25 (±4)
Lowstorey Vegetation (%)	0 (±0)	0 (±1)	10 (±7)	30 (±10)	20 (±4)	20 (±9)	40 (±9)	25 (±4)
Ground Vegetation Density (au)	46 (±1)	46 (±2)	44 (±1)	43 (±2)	35 (±3)	47 (±1)	38 (±2)	42 (±1)
Ground Vegetation Depth (cm)	54.6 (±5.56)	38.0 (±14.5)	33.1 (±9.09)	64.8 (±25.9)	90.7 (±2.81)	35.1 (±7.49)	83.0 (±15.5)	74.4 (±12.4)
Canopy								
Canopy Height (m)	9 (±1)	13 (±2)	15 (±1)	15 (±1)	20 (±4)	18 (±1)	23 (±2)	22 (±2)
Canopy Cover (%)	4.38 (±0.55)	26.19 (±12.4)	68.97 (±9.52)	62.24 (±10.20)	82.59 (±1.24)	76.69 (±2.37)	84.30 (±3.59)	82.25 (±1.55)
Tree Density (per 100 m^2)								
Sapling	0.83 (±0.15)	3.74 (±0.87)	3.68 (±1.35)	6.84 (±3.59)	46.4 (±14.3)	8.65 (±4.67)	32.3 (±11)	15.0 (±3.31)
Small Tree	-	2.01 (±0.77)	5.52 (±2.04)	2.14 (±0.75)	31.7 (±9.04)	2.5 (±0.75)	11.2 (±2.76)	20 (±2.89)
Large Tree	0.553 (±0.03)	1.93 (±1.08)	6.56 (±4.98)	2.18 (±0.80)	4.49 (±0.60)	8.12 (±1.52)	3.99 (±1.36)	13.1 (±3.52)
Climate								
Temperature (°C)	33.93 (±0.66)	34.70 (±0.76)	30.25 (±0.48)	31.44 (±0.80)	30.13 (±0.37)	29.17 (±0.42)	29.65 (±0.22)	29.93 (±0.44)
Humidity (%rh)	73.68 (±1.04)	71.35 (±1.61)	82.46 (±1.49)	77.90 (±1.61)	80.88 (±0.40)	81.25 (±0.94)	83.25 (±1.04)	81.18 (±0.89)



Appendix Figure 3. *Summary boxplots of environmental characteristics at each site.* **A & B)** Temperature and humidity were recorded for every sampling effort. **C & D**) Estimated percentage ground vegetation cover (0 – 2m height above ground) and lowstorey vegetation cover (2 – 5m height above ground) to the nearest 5%. **E)** Ground vegetation density on a scale of 0 to 50 arbitrary units (a.u.), a greater value equals a lower vegetation density (a maximum value of 50 indicates a relative vegetation density of 0%). **F)** Depth of ground vegetation. **G)** Canopy height. **H)** Percentage canopy cover. **I, J & K)** Log10 plots of the density of saplings (DBH < 0.05m), small trees (DBH 0.05 – 0.15m) and large trees (DBH > 0.15m) per m^2 respectively. (* denotes average)

Appendix table 3. *GPS Coordinates of each transect.*

Transect ID	GPS Coordinates (Longtitude, Latitude)	
	Start Point	End Point
KSA	117.984718, 5.417790	117.983040, 5.417122
KSR	117.987117, 5.413695	117.985308, 5.413493
KSN	117.980508, 5.413235	117.978901, 5.412386
KLA1	117.966838, 5.421197	117.967652, 5.420768
KLA2	117.966985, 5.421391	117.967797, 5.420982
KLR1	117.969667, 5.425253	117.969550, 5.424366
KLR2	117.969953, 5.425263	117.969923, 5.424363
KLN1	117.962906, 5.420650	117.962395, 5.419960
KLN2	117.963401, 5.420080	117.962559, 5.419721
Hillco	N5° 25.260' E118° 01.627'	N5° 25.263' E118° 01.520'
Pendirosa	N5° 24.733' E118° 06.025'	N5° 24.731' E118° 06.133'