

# A Network-Based Method to Detect Patterns of Local Crop Biodiversity: Validation at the Species and Infra-Species Levels

Mathieu Thomas<sup>\*,†,‡,1</sup>, Nicolas Verzelen<sup>§</sup>, Pierre Barbillon<sup>¶</sup>,  
 Oliver T. Coomes<sup>||</sup>, Sophie Caillon<sup>†</sup>, Doyle McKey<sup>†,#</sup>, Marianne Elias<sup>\*\*</sup>,  
 Eric Garine<sup>††</sup>, Christine Raimond<sup>‡‡</sup>, Edmond Dounias<sup>†</sup>,  
 Devra Jarvis<sup>§§,¶¶</sup>, Jean Wencélius<sup>††</sup>, Christian Leclerc<sup>|||</sup>,  
 Vanesse Labeyrie<sup>|||</sup>, Pham Hung Cuong<sup>##</sup>, Nguyen Thi Ngoc Hue<sup>##</sup>,  
 Bhuwon Sthapit<sup>\*\*\*</sup>, Ram Bahadur Rana<sup>†††</sup>, Adeline Barnaud<sup>‡‡‡,§§§,¶¶¶</sup>,  
 Chloé Violon<sup>††</sup>, Luis Manuel Arias Reyes<sup>|||</sup>,  
 Luis Latournerie Moreno<sup>###</sup>, Paola De Santis<sup>§§</sup>, François Massol<sup>\*\*\*\*</sup>

\*INRA, UMR 0320/UMR 8120 Génétique Quantitative et Évolution-Le Moulon, Gif-sur-Yvette, France

†CEFE UMR 5175, CNRS-Université de Montpellier, Université Paul-Valéry Montpellier, EPHE, Montpellier, France

‡FRB, CESAB (Centre de synthèse et d'analyse de la biodiversité), Technopôle de l'Environnement Arbois-Méditerranée, Aix-en-Provence, France

§UMR 729 Mathématiques, Informatique et STatistique pour l'Environnement et l'Agronomie, INRA SUPAGRO, Montpellier, France

¶AgroParisTech/UMR INRA MIA, Paris, France

||Department of Geography, McGill University, Montreal, Quebec Canada

#Institut Universitaire de France, Paris, France

\*\*Institut de Systématique, Évolution, Biodiversité ISYEB-UMR 7205-CNRS, MNHN, UPMC, EPHE Muséum national d'Histoire naturelle, Sorbonne Universités, Paris, France

††Université Paris Ouest/CNRS, UMR 7186 LESC, Nanterre, France

‡‡CNRS-UMR 8586 PRODIG, Paris, France

§§Bioversity International, Rome, Italy

¶¶Department of Crop and Soil Sciences, Washington State University, Pullman, Washington, USA

|||CIRAD, UMR AGAP, Montpellier, France

##Plant Resources Center, VAAS, MARD, Hanoi, Viet Nam

\*\*\*Bioversity International, Office of Nepal, Pokhara City, Nepal

†††Local Initiatives for Biodiversity, Research and Development (LI-BIRD), Pokhara, Kaski, Nepal

‡‡‡IRD, UMR DIADE, Montpellier, France

§§§LMI LAPSE, Dakar, Senegal

¶¶¶ISRA, LNRPV, Centre de Bel Air, Dakar, Senegal

|||CINVESTAV-IPN Unidad Mérida, Mérida, Yucatan, Mexico

###Instituto Tecnológico de Conkal, Division de Estudios de Posgrado e Investigación Conkal, Conkal, Yucatan, Mexico

\*\*\*\*Unité Evolution, Ecologie et Paléontologie (EEP), CNRS UMR 8198, Université Lille, Lille, France

<sup>1</sup>Corresponding author: e-mail address: thomas@moulon.inra.fr

## Contents

1. Introduction	261
2. Description of the Datasets Used in the Meta-Analysis	266
3. Description of the Methodological Framework	273
3.1 Mathematical Formalism	274
3.2 Variability of Farms' and Crops' Degrees	275
3.3 Revealing Data Structure Through LBMs	278
3.4 Uncovering Outliers Through PCA	281
3.5 Measuring Originality of Farms' Contributions Through Diversity Measures	286
4. Patterns of Local Crop Diversity: Results of the Meta-Analysis	292
4.1 Variability of Farms' and Crops' Degrees	292
4.2 Structure Detection Through Model-Based Clustering (LBM)	297
4.3 Outlier Detection Through PCA	298
4.4 Farms' Contributions to Local Diversity	298
5. Discussion	300
5.1 Contrasted Patterns of Local Crop Diversity at the Species and Infra-Species Levels	300
5.2 Relevance of Network-Based Methods	304
6. Conclusion	306
Acknowledgements	307
LBM Representation	308
Outlier Representation	311
Statistical Power Study of the Test Measuring the Impact of Crop-Rich and Crop-Poor Farms	312
Estimation of Nestedness	314
Glossary	316
References	316

## Abstract

In this chapter, we develop new indicators and statistical tests to characterize patterns of crop diversity at local scales to better understand interactions between ecological and socio-cultural functions of agroecosystems. Farms, where a large number of crops (species or landraces) is grown, are known to contribute a large part of the locally available diversity of both rare and common crops but the role of farms with low diversity remains little understood: do they grow only common varieties—following a nestedness pattern typical of mutualistic networks in ecology—or do ‘crop-poor’ farmers also grow rare varieties? This question is pivotal in ongoing efforts to assess the local-scale contribution of small farms to global agrobiodiversity. We develop new network-based approaches to characterize the distribution of local crop diversity (species and infra-species) at the village level and to validate these approaches using meta-datasets from 10 countries. Our results highlight the sources of heterogeneity in crop diversity at the village level. We often identify two or more groups of farms based on their different levels of diversity. In some datasets, ‘crop-poor’ farms significantly contribute to the local crop diversity. Generally, we find that the distribution of crop diversity is more heterogeneous at

the species than at the infra-species level. This analysis reveals the absence of a general pattern of crop diversity distribution, suggesting strong dependence on local agro-ecological and socio-cultural contexts. These different patterns of crop diversity distribution reflect an heterogeneity in farmers' self-organized action in cultivating and maintaining local crop diversity, which ensures the adaptability of agroecosystems to global change.



## 1. INTRODUCTION

Agriculture relies on the use of crop plant species to provision human societies with food, clothing, medicinal and narcotic substances, fodder for domestic animals, building materials and more recently with biofuel. Plant crop species were domesticated from wild ancestors, which often display variability in traits adapted to the local environment. During domestication, only a subset of diversity from the wild ancestors was selected, and shaped by the goals of farmers to produce a diversity of landraces, named and managed as distinct entities (Diamond, 2002). Furthermore, different crop species play distinctive, often complementary, roles in agriculture. For instance, including legumes in rotations or in associations with cereals limit the use of external inputs of fertilizer by increasing nutrient inputs through nitrogen fixation (Drinkwater et al, 1998). In many agroecosystems, the end result of these processes of selection among wild diversity, i.e., selection in farmers' fields and adoption of numerous kinds of crops, is a substantial increase in the diversity of cultivated plants, both in terms of the number of species and landrace diversity within species (Jarvis et al., 2008). Modernization of agriculture in industrialized countries during the twentieth century increased agricultural productivity thanks to uniformization, i.e., reduction of the number of crop species and varieties and genetic homogenization of varieties (Bonneuil et al., 2012). This genetic erosion was accompanied by the disruption of interactions among crop and wild species (Macfadyen and Bohan, 2010). This strategy also required an intensive use of fertilizers, pesticides, water and fossil fuels, creating strong environmental perturbations, including habitat fragmentation, soil erosion, water pollution, causing great reduction of wild biodiversity (MEA, 2005). Ecological, economic and social consequences of intensive agriculture are now identified and formalized by the Millennium Ecosystem Assessment. Recommendations for limiting these treats rely on an ecosystemic and transdisciplinary approach to the problem (Mulder et al., this issue). Central to this approach is identifying

trade-offs and synergies among ecological functional groups within ecosystems, and estimating their impacts on provisioning, regulating, supporting and cultural services. To maintain ecological synergies, manage trade-offs and sustain ecosystem resilience, values are assigned for these services and stakeholders are influenced through recognition, incentives and rewards based on these values (see [Butterfield et al., 2015](#), for more details). The same conceptual framework is applied to manage agroecosystems (agro-ecology), considering trade-offs between agroecosystem functions (pollination, ecological pest control, etc.) and provision of goods ([MEA, 2005](#)). This agro-ecological approach to agriculture is not fully satisfying because it often neglects social and cultural processes directly linked to agriculture, such as local knowledge concerning farming practices, which can make important contribution to ensuring the sustainability of agroecosystems ([Jackson et al., 2007](#); [Martin et al., 2010](#)). In this chapter, we consider agricultural systems as socio-agroecosystem in which social and cultural functions need to be examined in addition to ecological functions. More attention must be paid to farmers' practices in highly diversified systems, because these practices play a role in creating and maintaining diversity, which is often of great importance to system functioning. Understanding the interactions of these practices with biological and ecological processes is necessary to improve our understanding and management of synergies and trade-offs occurring in agroecosystems.

A primary requisite for understanding and predicting the sustainability of agroecosystems facing environmental, political, social and economic changes is to assess how these systems manage crop diversity (e.g. [Samberg et al., 2013](#)). For instance, in the case of manioc cultivated by the Makushi Amerindians of Guyana, some varieties are specially grown for particular dishes, some grow quickly thereby ensuring early yield, while still others grow slowly and act as an 'ever-present' insurance resource ([Elias et al., 2000](#)). Often, diversity is simply valued for its own sake ([Boster, 1985](#)), or as a means to foster social relations ([Empeaire and Peroni, 2007](#); [Heckler and Zent, 2008](#)). Another example of crop biodiversity maintenance is the great diversity of landraces present in the milpas of Meso-America, which are the end product of several 1000 years of directed selection on maize, beans, squash and chilli peppers by the farmers of the region ([Tuxill et al., 2010](#)). Understanding relationships among landraces makes it possible to gain insight into the cultural history. The particular traits exhibited by local varieties grown in milpas today reflect Yucatan farmers' short- and long-term responses to agro-environmental conditions, the ecological

demands of crop production and the aesthetic, culinary and religious sensibilities of farmers (Tuxill et al., 2010). Farming practices that maintain crop diversity are of paramount importance in helping crops and farmers adapt to global changes, notably climate change (Vigouroux et al., 2011) and the increasingly rapid emergence of agricultural pests (Diamond, 2002). In addition, cultivating diverse crops and varieties at the landscape level favours ecological and economic sustainability by reducing the need for chemical inputs (Bianchi et al., 2006; Crowder et al., 2010). Crop diversity also provides an insurance value. Although some combinations of species or varieties may be functionally redundant in an agroecosystem, at least at a given time, a subset of species and varieties may confer to the system the capacity to adapt to environmental fluctuations (Di Falco and Perrings, 2005; Jackson et al., 2007; Smale et al., 1998).

The spatial distribution of crop diversity is expected to be partially explained by environmental factors, due to the differential adaptation of crops to local conditions (Mariac et al., 2011). For instance, crops require different physiological adaptations to cope with limiting factors associated with dry and wet climates. Moreover, selective pressures in cultivated environments differ from those in wild environments. However, unless massive inputs liberate crops from environmental constraints, adaptation to local abiotic environments is expected to shape crop diversity—as it shapes the diversity of wild plants—at more or less large spatial scales, over latitudinal or elevational gradients (Vigouroux et al., 2011). At fine spatial scales, local adaptation is also expected to play a role in the distribution of crop diversity, due to the heterogeneity of soil quality of agricultural fields and to variability in local rainfall (Fraser et al., 2012).

In addition to environmental factors, it has been argued that crop diversity can only be understood if social and cultural aspects of the contextual environment are taken into account (Leclerc and Coppens d'Éeckenbrugge, 2012; Rival and McKey, 2008). Agricultural societies have shaped the diversity of their cultivated crops in ways that fit their traditions, habits, myths, social organizations and livelihoods (Delètre et al., 2011; Labeyrie et al., 2013). Indeed, crops and humans have likely evolved together, as cultural practices may have been shaped by available edible plants and agricultural selection may have answered cultural needs. The study of crop genetic and inter-specific diversity in the context of both environment- and society-driven selective pressures is now taken into account through the  $G \times E \times S$  framework (Leclerc and Coppens d'Éeckenbrugge, 2012). Thus, studying the distribution of crop diversity

and linking it with both social and environmental factors cannot be based on a solely biological perspective. However, interdisciplinary studies of the distribution of crop diversity must retain quantitative rigour and thus be based on a sound statistical framework. To date, the distribution of crop diversity has been assessed mostly through the use of diversity indices adopted from ecology and economics, indices of richness, evenness, concentration, etc. (e.g. [Jarvis et al., 2008](#)). These indices only make use of crop diversity data as an instance of ‘type in location’ data and this limits the questions that can be addressed. They can help explain why crops are found in the fields they are in, but not why farmers decided to cultivate a given crop, for example. We failed to find any studies that even came near to exploiting the potential of analyses of the network<sup>\*1</sup> feature of crop-by-farms datasets which includes social aspects, such as farmer-to-farmer circulation of seeds (and other propagules) of varieties and crop species. These bipartite networks\* are composed of two kinds of nodes\* representing a farm or a crop (species or landrace); an edge\* connects two kinds of nodes and means that a particular crop is grown in a particular farm.

Our main goal in this chapter is to answer the question ‘which farms contribute, and how, to the diversity of crops grown in a given village?’ by examining inventories of crops species and landraces grown at the farm level. To do this, we offer a novel methodological framework using network-based and null model-based statistical tests. From a methodological perspective, inventory datasets can be construed as bipartite networks, namely crop-by-farm interaction networks\*, in the same way as plant–pollinator or host–parasite interaction networks in ecology. In social network analysis, network approaches have been used to assess the properties of network processes linked to social institutions, such as friendship, advice or seed exchange networks (‘who interacts with whom’ or ‘who gives to whom’) ([Lazega et al., 2012](#); [Reyes-García et al., 2013](#); [Wasserman and Faust, 1994](#)). In ecology, on the other hand, networks have been used to study both contact networks (metapopulations or metacommunities) and structured interaction networks\*, such as food webs (herbivore–host plant networks) or mutualistic networks (plant–pollinator networks). When interaction partners can be clearly categorized (plants, pollinators; plants, herbivores and parasitoids), the use of bi- or multi-partite networks is an appropriate approach. In the present study, we develop a framework for the study of crop-by-farm datasets that makes use of the bipartite nature

---

<sup>1</sup> \* indicates that the word or expression is defined in Glossary section.

of the data to reveal potential patterns in the structure of diversity at the scale of the village or the clusters of interacting villages.

Our paper offers an alternative to the nestedness\* approach, for several reasons that are detailed below. The study of bipartite networks in ecology is a recent endeavour (Jordano, 1987). Over the past three decades, the topological properties of bipartite networks have been studied to answer a variety of questions, such as whether the networks are stable, robust to species extinctions or additions, functionally redundant, etc. (Astegiano et al., 2015; Gil et al., 2015; Jordano et al., 2003; Thébault and Fontaine, 2010). In particular, the nestedness of mutualistic bipartite networks has often been investigated and studies suggest that nestedness may be the key property explaining the dynamics and structural stability of mutualistic networks (Thébault and Fontaine, 2010). In spatial ecology studies, nested patterns are often explained as resulting from source–sink processes wherein species–rich locations function as sources producing many emigrating individuals which, in turn, contribute to the diversity in species–poor, sink locations (Atmar and Patterson, 1993). In mutualistic interaction networks, nestedness can be understood as arising from feasibility constraints on the existence of specialist–specialist interactions, i.e., nestedness decreases effective inter-specific competition and thus increases the number of species that can coexist (Bascompte and Jordano, 2007; Bastolla et al., 2009). In systems involving social as well as ecological processes, such as in the present case of crop–by–farm interactions, one may ask whether this nestedness pattern holds, as crops present in less diverse (crop–poor) farms could comprise a subset of those cultivated in more diverse farms. Among the Duupa in northern Cameroon, for example, older farmers accumulate crop diversity during their life (sources) and become sources of diversity for young farmers (sinks) (Alvarez et al., 2005). When crops are actively cultivated by farmers, for example, as staple food, copying other farmers’ portfolios of crops might result in strong similarities in cultivated diversity among fields, but not necessarily following a nested pattern. Therefore, contrary to the case for ecological systems, certain mechanistic reasons may justify considering crop–by–farm interactions as systematically nested, precluding explanations solely based on source–sink processes.

From a purely methodological perspective, the available indices of network nestedness are inconsistent, both in the value of nestedness metrics and in their associated  $p$ -value when confronted with the configuration model\*, a null model of partner interactions constrained by degree\*, i.e., fixing the degree of rows and columns (Podani and Schmera, 2012). Although

nestedness remains a largely verbal concept and its mathematical definition is in need of refinement, being able to detect nested patterns in crop diversity would be useful for characterizing the diversity of strategies used by farmers to cope with different socio-cultural and environmental contexts.

In [Section 2](#), we introduce a meta-dataset of specific and infra-specific crop diversity at the local scale in different agricultural contexts. In [Section 3](#), we describe our methodological framework, graphical representations when appropriate and the tests proposed, illustrated using ‘toy’, hypothetical examples. Our approach allows us: (i) to test whether the variability in the number of connections per farm and per crop type is different from random expectations under a homogeneous random graph model\* (Erdős-Rényi model\*); (ii) to reveal structure (modules, cores, etc.) in the dataset using latent-block models\* (LBMs); (iii) to uncover ‘outliers’ (farmers or crop types that do not conform to the general connection pattern) using principal component analyses (PCAs); and (iv) to measure and test the originality of farmers’ contributions to overall crop diversity using beta-diversity indices. In [Section 4](#), we perform a meta-analysis applying the methodological framework to our meta-dataset, which allows us to highlight both regularities and particularities among the datasets. Our approach yields graphical representations of the different tests (reordering of interactions in the case of LBMs or principal plane representations for PCAs) and non-parametric tests of our hypotheses, the significance of which is assessed through comparison with a permutation-based null model (the configuration model for graphs with given degrees). These graphical and statistical approaches are designed to be transferable to other similar problems in ecology. In [Section 5](#), we discuss our results and the value and the limits of our approach.



## 2. DESCRIPTION OF THE DATASETS USED IN THE META-ANALYSIS

Fifty published or unpublished datasets dealing with crop inventories were provided by ethnobiologists, geographers and ecologists for analysis ([Tables 1 and 2](#)). These data were collected in 10 different countries ([Fig. 1](#)) between 1998 and 2013. For each dataset, a partial set or the full set of farms from the same village was characterized for one of the two classes of operational taxonomic units (OTU) considered: the species or the infra-species (landrace) level. These data were gathered through direct interviews with the plant crop cultivators in the farm, a subset of them or only with the head of the farm. Datasets were retained when the number of characterized farms and the number



**Table 1** Description of the 18 Datasets Dealing with Specific Diversity (OTU = Species)

<b>Dataset</b>	<b>Country</b>	<b>Community</b>	<b>Village</b>	<b>Farm Sample Size</b>	<b>Crop Sample Size</b>	<b>Collect Year</b>	<b>Original Article</b>
CL-M01	Kenya	Tharaka	Kamarandi	95	16	2010	<a href="#">Labeysrie et al. (2013)</a>
CV-M01	Cameroon	Tupuri	Gulurgu-Lokoro	15	23	2011	Unpublished data
EG-M05	Cameroon	Duupa farmers	Ninga	14	58	2002	<a href="#">Gariné and Raimond (2005)</a>
EG-M08	Cameroon	Duupa farmers	Wante	18	68	2002	<a href="#">Gariné and Raimond (2005)</a>
OC-M02	Peru	Corrientes River	Boca del Copal	19	108	2003	<a href="#">Perrault-Archambault and Coomes (2008)</a>
OC-M04	Peru	Corrientes River	San Juan de Trompeteros	35	120	2003	<a href="#">Perrault-Archambault and Coomes (2008)</a>
OC-M05	Peru	Corrientes River	San Juan de Trompetero Nativo	22	108	2003	<a href="#">Perrault-Archambault and Coomes (2008)</a>
OC-M06	Peru	Corrientes River	San Jose de Porvenir	18	84	2003	<a href="#">Perrault-Archambault and Coomes (2008)</a>
OC-M07	Peru	Corrientes River	Nuevo Porvenir	22	83	2003	<a href="#">Perrault-Archambault and Coomes (2008)</a>
OC-M09	Peru	Corrientes River	Nuevo Paraiso	11	83	2003	<a href="#">Perrault-Archambault and Coomes (2008)</a>
OC-M10	Peru	Corrientes River	Nuevo Peruanito	15	88	2003	<a href="#">Perrault-Archambault and Coomes (2008)</a>

*Continued*

**Table 1** Description of the 18 Datasets Dealing with Specific Diversity (OTU = Species)—cont'd

<b>Dataset</b>	<b>Country</b>	<b>Community</b>	<b>Village</b>	<b>Farm Sample Size</b>	<b>Crop Sample Size</b>	<b>Collect Year</b>	<b>Original Article</b>
OC-M11	Peru	Corrientes River	Nuevo Pucacuro	54	161	2003	<a href="#">Perrault-Archambault and Coomes (2008)</a>
OC-M12	Peru	Corrientes River	Santa Rosa	14	124	2003	<a href="#">Perrault-Archambault and Coomes (2008)</a>
OC-M13	Peru	Corrientes River	Santa Elena	30	153	2003	<a href="#">Perrault-Archambault and Coomes (2008)</a>
OC-M14	Peru	Corrientes River	San Jose de Nueva Esperanza	24	139	2003	<a href="#">Perrault-Archambault and Coomes (2008)</a>
OC-M16	Peru	Corrientes River	Valencia	21	147	2003	<a href="#">Perrault-Archambault and Coomes (2008)</a>
SC-M05	Vanuatu	Vanua Lava, Banks group	Eastern coast	15	37	2007– 2009	Unpublished data
SC-M06	Ecuador	Huaorani	Guiyero	13	15	2000	Unpublished data

**Table 2** Description of the 33 Datasets Dealing with Infra-Specific Diversity (OTU = Landrace)

Dataset	Country	Community	Village	Species	Predominant Propagation Mode	Farms Sample Size	Crop Sample Size	Collect Year	Original Article
AB-M02	Cameroon	Duupa farmers	Wante	Sorghum ( <i>Sorghum bicolor</i> )	Partially outcrossing	13	23	2003	Unpublished data
CL-M02	Kenya	Tharaka	Kamarandi	Sorghum ( <i>Sorghum bicolor</i> )	Partially outcrossing	95	20	2010	<a href="#">Labeyrie et al. (2013)</a>
CV-M02	Cameroon	Tupuri	Gulurgu-Lokoro	Sorghum ( <i>Sorghum bicolor</i> )	Partially outcrossing	15	22	2011	Unpublished data
DJ-M003a	Nepal	Kaski	village9	Rice ( <i>Oryza sativa</i> )	Inbreeding	33	24	2006	<a href="#">Jarvis et al. (2008)</a>
DJ-M003b	Nepal	Kaski	village10	Rice ( <i>Oryza sativa</i> )	Inbreeding	52	32	2006	<a href="#">Jarvis et al. (2008)</a>
DJ-M003c	Nepal	Kaski	village11	Rice ( <i>Oryza sativa</i> )	Inbreeding	24	21	2006	<a href="#">Jarvis et al. (2008)</a>
DJ-M003d	Nepal	Kaski	village14	Rice ( <i>Oryza sativa</i> )	Inbreeding	25	18	2006	<a href="#">Jarvis et al. (2008)</a>
DJ-M009a	Nepal	Bara	village1	Rice ( <i>Oryza sativa</i> )	Inbreeding	35	11	2006	<a href="#">Jarvis et al. (2008)</a>
DJ-M009b	Nepal	Bara	village2	Rice ( <i>Oryza sativa</i> )	Inbreeding	29	12	2006	<a href="#">Jarvis et al. (2008)</a>

*Continued*

**Table 2** Description of the 33 Datasets Dealing with Infra-Specific Diversity (OTU = Landrace)—cont'd

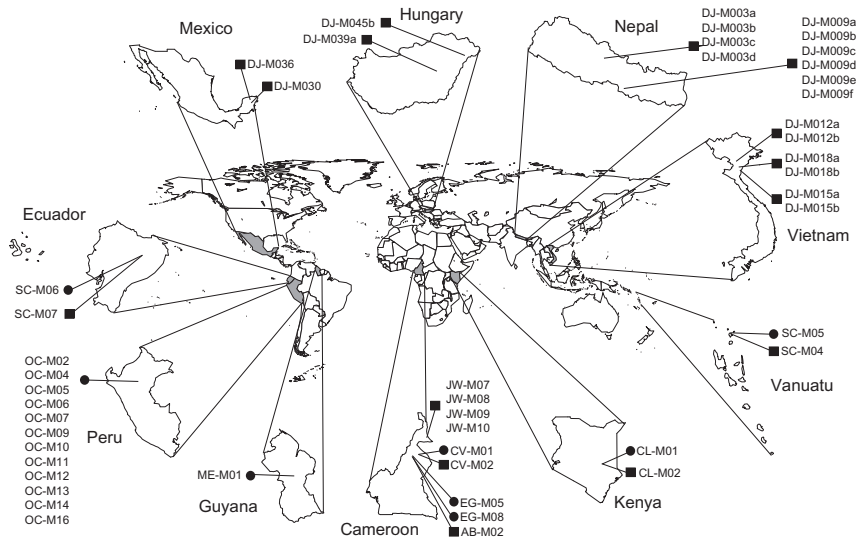
<b>Dataset</b>	<b>Country</b>	<b>Community</b>	<b>Village</b>	<b>Species</b>	<b>Predominant Propagation Mode</b>	<b>Farms Sample Size</b>	<b>Crop Sample Size</b>	<b>Collect Year</b>	<b>Original Article</b>
DJ-M009c	Nepal	Bara	village3	Rice ( <i>Oryza sativa</i> )	Inbreeding	37	14	2006	Jarvis et al. (2008)
DJ-M009d	Nepal	Bara	village4	Rice ( <i>Oryza sativa</i> )	Inbreeding	14	8	2006	Jarvis et al. (2008)
DJ-M009e	Nepal	Bara	village5	Rice ( <i>Oryza sativa</i> )	Inbreeding	31	14	2006	Jarvis et al. (2008)
DJ-M009f	Nepal	Bara	village6	Rice ( <i>Oryza sativa</i> )	Inbreeding	29	16	2006	Jarvis et al. (2008)
DJ-M012a	Vietnam	Dabac	Cang	Rice ( <i>Oryza sativa</i> )	Inbreeding	58	42		Jarvis et al. (2008)
DJ-M012b	Vietnam	Dabac	Tat	Rice ( <i>Oryza sativa</i> )	Inbreeding	57	58		Jarvis et al. (2008)
DJ-M015a	Vietnam	Nghiahung	Dong Lac	Rice ( <i>Oryza sativa</i> )	Inbreeding	58	42		Jarvis et al. (2008)
DJ-M015b	Vietnam	Nghiahung	Kien Thanh	Rice ( <i>Oryza sativa</i> )	Inbreeding	57	58		Jarvis et al. (2008)
DJ-M018a	Vietnam	Nhoquan	Quang Mao	Rice ( <i>Oryza sativa</i> )	Inbreeding	58	42		Jarvis et al. (2008)

DJ-M018b	Vietnam	Nhoquan	Yen Minh	Rice ( <i>Oryza sativa</i> )	Inbreeding	57	58		Jarvis et al. (2008)
DJ-M030	Mexico	Ichmul	Multi-village	Maize ( <i>Zea mays</i> )	Outcrossing	101	11		Jarvis et al. (2008)
DJ-M036	Mexico	Yaxcaba	Yaxcaba	Maize ( <i>Zea mays</i> )	Outcrossing	61	13	1999	Jarvis et al. (2008)
DJ-M039a	Hungary	Dévaványa	village1	Bean ( <i>Phaseolus vulgaris</i> , <i>Phaseolus lunatus</i> , <i>Vigna unguiculata</i> )	Inbreeding	13	10		Jarvis et al. (2008)
DJ-M045b	Hungary	Szatmár-Bereg	village2	Bean ( <i>Phaseolus vulgaris</i> , <i>Phaseolus lunatus</i> , <i>Vigna unguiculata</i> )	Inbreeding	18	12		Jarvis et al. (2008)
DJ-M045c	Hungary	Szatmár-Bereg	village3	Bean ( <i>Phaseolus vulgaris</i> , <i>Phaseolus lunatus</i> , <i>Vigna unguiculata</i> )	Inbreeding	10	12		Jarvis et al. (2008)
DJ-M045d	Hungary	Szatmár-Bereg	village4	Bean ( <i>Phaseolus vulgaris</i> , <i>Phaseolus lunatus</i> , <i>Vigna unguiculata</i> )	Inbreeding	12	10		Jarvis et al. (2008)

Continued

**Table 2** Description of the 33 Datasets Dealing with Infra-Specific Diversity (OTU = Landrace)—cont'd

Dataset	Country	Community	Village	Species	Predominant Propagation Mode	Farms Sample Size	Crop Sample Size	Collect Year	Original Article
JW-M07	Cameroon		Nulda	Sorghum ( <i>Sorghum bicolor</i> )	Partially outcrossing	35	22	2008–2009	Unpublished data
JW-M08	Cameroon		Nulda	Sorghum ( <i>Sorghum bicolor</i> )	Partially outcrossing	45	24	2009–2010	Unpublished data
JW-M09	Cameroon		Nulda	Sorghum ( <i>Sorghum bicolor</i> )	Partially outcrossing	51	27	2011–2012	Unpublished data
JW-M10	Cameroon		Nulda	Sorghum ( <i>Sorghum bicolor</i> )	Partially outcrossing	15	21	2012–2013	Unpublished data
ME-M01	Guyana	Makushi Amerindians	Rewa	Cassava ( <i>Manihot esculenta</i> )	Clonal	24	75	1997–1998	<a href="#">Elias et al. (2000)</a>
SC-M04	Vanuatu	Vanua Lava, Banks group	Eastern coast	Taro ( <i>Colocasia esculenta</i> )	Clonal	15	34	2007–2009	Unpublished data
SC-M07	Ecuador	Huaorani	Guiyero	Manioc ( <i>Manihot esculenta</i> )	Clonal	13	29	2000	Unpublished data



**Figure 1** Map showing locations of the different datasets used in the meta-analysis. Filled circles correspond to the datasets collected at the specific level and filled squares correspond to the dataset collected at the infra-specific level.

of crops were both greater than 10. For 18 datasets, information was collected at the species level (Table 1); for 32 datasets, information was collected at the landrace level, which corresponds to the terminal taxon in the farmers' local naming systems, covering seven different species (maize, rice, wheat, bean, manioc, taro and sorghum) which correspond to the major crops of the areas under study (Table 2). These species are characterized by their predominant propagation mode (partially outcrossing, outcrossing, inbreeding and clonal) following the classification proposed by Jarvis et al. (2008). Data were structured following a rectangular incidence matrix\* with farms in rows and species or landrace in columns, and represented as a bipartite network. Data collected at the species or infra-species level represent two levels of local crop biodiversity. Underlying processes shaping the distribution of local crop diversity are assumed to be different for these two levels. Therefore, species and infra-species data are analysed and described separately.



### 3. DESCRIPTION OF THE METHODOLOGICAL FRAMEWORK

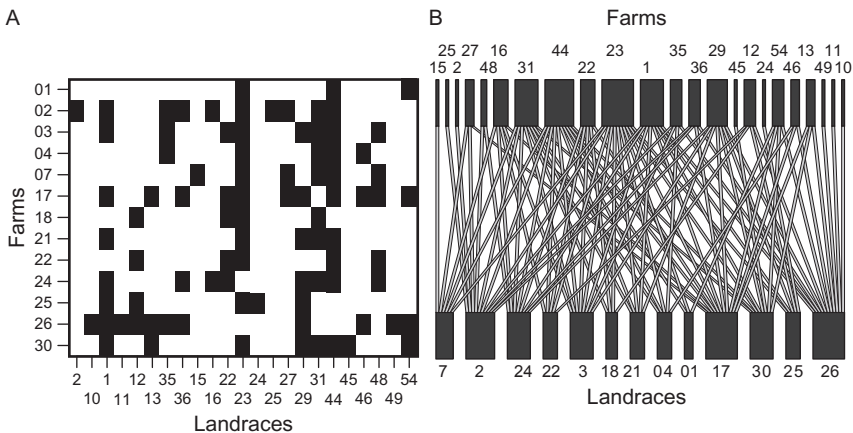
This section introduces the statistical framework for analysing crop-by-farm network data. After defining the main concepts, we detail the four

main steps of the analysis. First, the degree distribution of the data is evaluated as a way to test whether a completely random model (Erdős-Rényi model) fits well the data. Second, we use a LBM to investigate more thoroughly the structure of the network. This method pinpoints groups of farms and groups of crops that tend to be highly connected. Third, we test whether this high-level structure (blocks) is different from what can be expected simply by features of the low-level structure such as degree heterogeneity. The methods comprised by these two last steps provide new graphical representations of the network data emphasizing the studied patterns. Finally, complementary analyses based on diversity measures are introduced. In each subsection, toy examples illustrate the purpose, the benefits and the limitations of the proposed methods.

### 3.1 Mathematical Formalism

In the following, we denote  $n$  is the number of farms and  $m$  is the number of crops. The incidence matrix (with farms as rows and crops as columns) that summarizes the data is noted  $\mathbf{X}$ , so that  $X_{ij} = 1$  when farm  $i$  cultivates crop  $j$ . Using this representation (Fig. 2A), we can readily apply statistical methods for binary matrices.

Any incidence matrix can also be treated as the adjacency matrix of some bipartite graph  $G$ . More specifically, consider a collection of nodes corresponding to all farms and all crops (species or landraces) and put an edge



**Figure 2** (A) Example of an incidence matrix with farms in lines and crops in columns, and where 0 are black cells and 1 are white cells and (B) example of a crop-by-farm bipartite network between farms and landraces (dataset AB-M02).



between the farm  $i$  and the crop  $j$  if and only if  $X_{ij} = 1$ . The obtained network is bipartite (Fig. 2B) as no two farms and no two crops are connected in the network. Building on this equivalence between incidence matrices and bipartite graphs, we can borrow methodologies developed in the field of network analysis (Kolaczyk, 2009).

As these two representations are equivalent, any statistical analysis could be defined either in terms of the incidence matrix or in terms of the bipartite network  $G$ . For ease of reading, this chapter makes use of the incidence matrix terminology but we sometimes borrow network notations to emphasize the connection with the literature on network analysis.

Summing over crop, the number of crops cultivated on farm  $i$ ,  $C_i$ , is

$$C_i = \sum_j X_{ij}. \quad (1)$$

Summing over farms, the number of farms where crop  $j$  is cultivated,  $F_j$ , is

$$F_j = \sum_i X_{ij}. \quad (2)$$

Quantities  $N$ ,  $C_i$ ,  $F_j$  and  $X_{ij}$  are finally linked by the following relations:

$$N = \sum_i C_i = \sum_j F_j = \sum_{i,j} X_{ij}. \quad (3)$$

Following network terminology,  $C_i$  is also called the farm's degree and  $F_j$  the crop's degree.

## 3.2 Variability of Farms' and Crops' Degrees

### 3.2.1 Description of the Test on Degree Distributions

First, we evaluate whether all farms in the same village grow a similar number of crop or if there is high heterogeneity between farms' crop richness. Formally, we test whether the degrees  $F_j$  follow binomial distributions by considering a statistic  $T$  that compares the observed variance of the crops' degree with the one that would have been expected if the degrees  $C_i$  were following independent and identically distributed (*iid*) binomial distributions.

$$T_{\text{row}} := \frac{\widehat{\text{Var}}(C)}{n\hat{p}(1-\hat{p})},$$

where  $\hat{p} = N/nm$  is the density of the incidence matrix and  $\widehat{\text{Var}}(C) = 1/(n-1) \sum_{i=1}^n (C_i - m\hat{p})^2$  is the empirical variance of  $(C_i)$ ,

$i = 1, \dots, n$ . Large  $T_{\text{row}}$  values suggest that the farms' crop richness is highly heterogeneous, whereas small  $T_{\text{row}}$  values suggest more equity. The statistical significance of  $T$  is assessed by a parametric bootstrap method working as follows. For  $i = 1, \dots, n_{\text{sim}}$ , a new incidence matrix  $\mathbf{X}^{(i)}$  is generated by sampling independent Bernoulli distributions with parameters  $\hat{p}$  in each entry. For all these matrices, the link density  $\hat{p}^{(i)}$ , the empirical variance of the farms' degrees  $\widehat{\text{Var}}^{(i)}(C)$  and the variance ratio  $T_{\text{row}}^{(i)}$  are computed. Finally, the left and right  $p$ -values are, respectively,  $p\text{val}_{L,\text{row}} := \frac{\#\{i: T_{\text{row}}^{(i)} < T_{\text{row}}\}}{n}$  and  $p\text{val}_{R,\text{row}} := \frac{\#\{i: T_{\text{row}}^{(i)} > T_{\text{row}}\}}{n}$ .

The crops' degree distribution is evaluated in a similar fashion:

$$T_{\text{col}} := \frac{\widehat{\text{Var}}(F)}{m\hat{p}(1-\hat{p})};$$

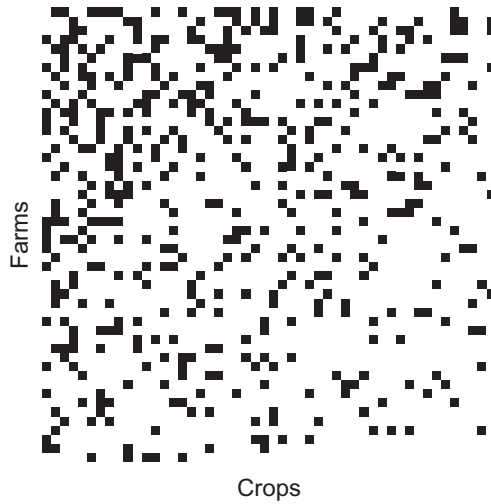
$$\widehat{\text{Var}}(F) = \frac{1}{m-1} \sum_{j=1}^m (F_j - n\hat{p})^2.$$

The corresponding  $p$ -values are also evaluated by parametric bootstrap. In our analysis, the parameter  $n_{\text{sim}}$  is fixed to 10,000.

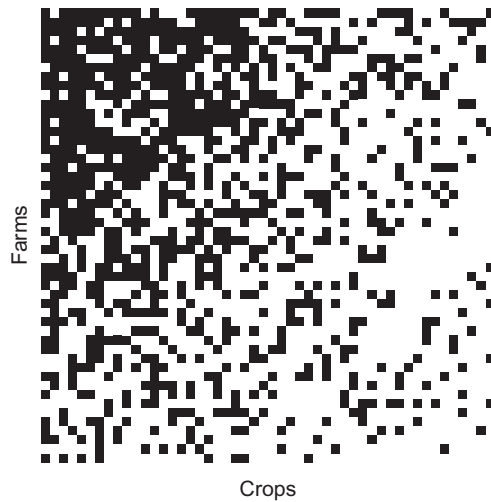
Under an Erdős-Rényi null model, where all the entries of  $\mathbf{X}$  follow independent Bernoulli distributions with identical parameters, and the farms' and crops' degrees follow binomial distributions. Consequently, any small  $p$ -value ( $p\text{val}_{L,\text{row}}, p\text{val}_{R,\text{row}}, p\text{val}_{L,\text{col}}, p\text{val}_{R,\text{col}}$ ) would indicate that this Erdős-Rényi model is not realistic.

### 3.2.2 Application of the Test on Degree Distributions to a Toy Example

Figures 3–5 display three examples of incidence matrices. The last two matrices were generated by organizing groups of crops and groups of farms according to a LBM (see presentation in the next subsection). The farms and the crops were sorted by degrees within groups. Note that this structure of groups is generally unknown in real datasets and has to be recovered by statistical inference techniques. In Fig. 3, the incidence matrix was generated from *iid*. Bernoulli random variables. Hence, its row and column degrees follow binomial distributions. This corresponds to the null hypothesis of the test on the variance of degrees. The tests are non-significant for this incidence matrix (Table 3). In Fig. 4, some farms were assumed to grow more crops than others and some crops were assumed to be more common than others. Therefore, as expected, the tests on the variance of degrees show clearly an over-dispersion for farms and crops. In Fig. 5, there exist particular associations between some groups of farms



**Figure 3** Incidence matrix with entries generated independently and identically distributed according to a Bernoulli distribution with probability 0.2.



**Figure 4** Incidence matrix generated with heterogeneous distribution for different groups of crops and farms (see Fig. 7 in next subsection for details). Some farms grow more crops than other and some crops are more common than others.

and some groups of crops although the degree is quite homogeneous for farms; crops heterogeneity appears because the groups of farms are not of the same size.

As illustrated in these three examples, the tests on the variance of degrees may detect heterogeneity but some particular structure of association may be



**Figure 5** Incidence matrix generated with distribution implying particular association between crops and farms (see Fig. 6 in next subsection for details). Two groups of crops are mainly grown by corresponding subgroups of farms.

**Table 3**  $p$ -Values for Tests on the Variability of Degrees for Farms and Crops (Left: Under-Dispersion, Right: Over-Dispersion) Applied on the Three Toy Examples Presented in Figs. 3–5

	Farms		Crops	
	Left	Right	Left	Right
Fig. 3	0.8143	0.1857	0.6345	0.3655
Fig. 4	1	<0.001	1	<0.001
Fig. 5	0.1604	0.8396	0.9924	0.0076

missed as in the case of Fig. 5. Indeed, the tests are performed independently on farms and on crops and thus are not able to detect patterns of association.

### 3.3 Revealing Data Structure Through LBMs

#### 3.3.1 Description of the LBMs

In order to cluster the farms and the crops simultaneously on the basis of the incidence matrix  $\mathbf{X}$ , we propose to use a probabilistic model called LBM (Govaert and Nadif, 2008; Keribin et al., 2014), which assumes a mixture distribution both on the farms and crops. According to this model, the network is generated relying on latent blocks (also called clusters) of farms and

of crops. The probability that a crop  $j$  is grown on a farm  $i$  is conditioned to these latent blocks and depends only on the block  $V(i)$  to which farm  $i$  belongs and the block  $W_j$  to which crop  $j$  belongs. For all  $1 \leq i \leq n$ ,  $1 \leq j \leq m$ ,  $1 \leq q \leq Q$  and  $1 \leq l \leq L$ , the probability that  $i$  belongs to block  $q$ , that  $j$  belongs to block  $l$  and the conditional probability of  $X_{ij}$  given the block  $V_i$  and  $W_j$  are, respectively, denoted

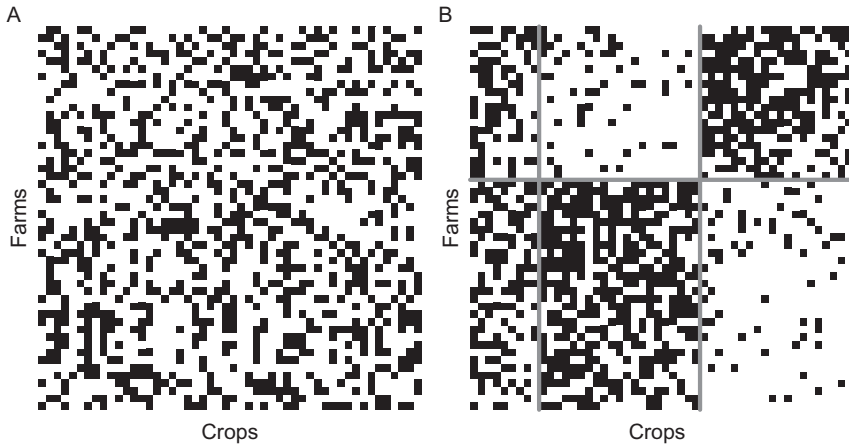
$$\begin{aligned}\mathbb{P}(V_i = q) &= \alpha_q, \\ \mathbb{P}(W_j = l) &= \beta_l, \\ \mathbb{P}(X_{ij} = 1 | V_i = q, W_j = l) &= \pi_{ql},\end{aligned}$$

where  $\theta = (\alpha_1, \dots, \alpha_Q, \beta_1, \dots, \beta_L, \pi_{11}, \dots, \pi_{QL})$  is the vector of unknown parameters to be estimated under the obvious constraints  $\sum_q \alpha_q = 1$ ,  $\sum_l \beta_l = 1$ . This model is quite flexible because it can account not only for situations where there is modularity, i.e., a unique block of crops is associated with each block of farms and these farms tend to grow mainly crops from that block and few from other blocks, but also for situations where there are richer farms (growing significantly more crops than others) and/or more common crops (grown by significantly more farms than others).

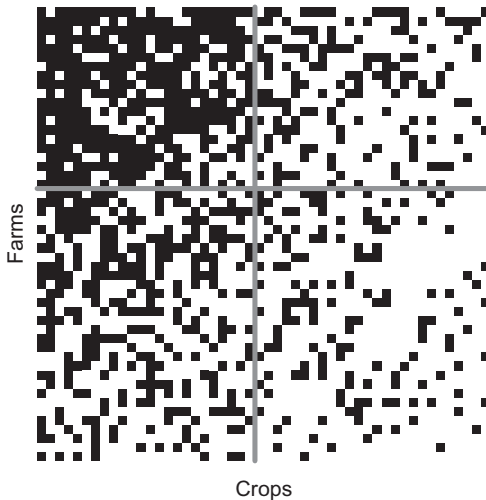
The standard procedures to obtain maximum likelihood estimates when dealing with latent variables rely on the expectation–maximization (EM) algorithm (Dempster et al., 1977). However, the computation of the conditional distribution of the latent variables with respect to the observed data is not tractable, which makes the  $E$ -step unfeasible. Following Govaert and Nadif (2008), we use a variational approach to cope with this difficulty. The number of blocks of farms  $Q$  and the number of blocks of crops  $L$  are chosen thanks to the integrated completed likelihood (ICL) criterion as proposed by Keribin et al. (2014). Once the parameters have been estimated, we obtain as a by-product the posterior probabilities  $\mathbb{P}(V_i = q | \mathbf{X})$  and  $\mathbb{P}(W_i = l | \mathbf{X})$ , from which the true blocks are estimated. We can then provide a new representation of the incidence matrix  $\mathbf{X}$  where the rows (farms) and the columns (crops) have been re-organized in homogeneous blocks. We used the R package `blockmodels` (Leger, 2015) to perform the estimations and the model selection.

### 3.3.2 Application of LBM to a Toy Example

Figures 6–8 are illustrations of the block clustering provided by the LBM in three typical cases. The cases of Figs. 6 and 7 are the same as those in Figs. 5 and 4, respectively. The groups were considered as latent/unknown and the

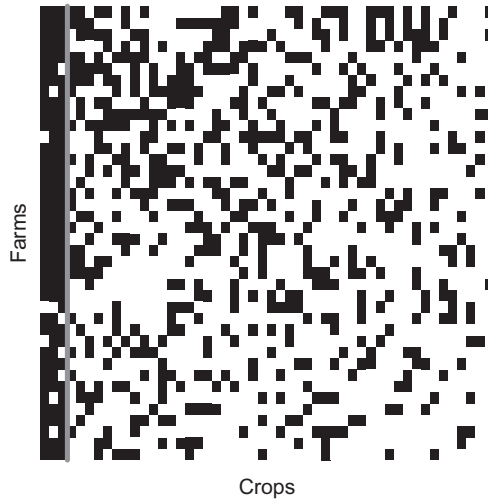


**Figure 6** Incidence matrix generated according to a LBM with three blocks of crops, two blocks of farms and  $\pi = (0.5 \ 0.1 \ 0.6 \ 0.5 \ 0.6 \ 0.1)$ . (A) Observed incidence matrix and (B) same incidence matrix re-organized and clustered in homogeneous blocks obtained by LBM inference.



**Figure 7** LBM clustering when the data are generated with two blocks of farms (rich and poor farms), two blocks of crops (rare and common crops) and  $\pi = (0.7 \ 0.3 \ 0.4 \ 0.2)$ .

farms and crops were clustered in homogeneous blocks by using the inference procedure described above. This is illustrated in Fig. 6, where the same incidence matrix is plotted before and after re-organization according to the estimated blocks. In Fig. 6, the difference between the two groups of farms comes from the two last groups of crops. The first group of crops is equally



**Figure 8** LBM clustering when the data are generated with one block of farms, two blocks of crops (one block with only three crops) and  $\pi = (0.9 \ 0.3)$ .

grown on farms of any group. In contrast, the second group of crops is mainly grown by the second group of farms and the third group of crops is mainly grown by the first group of farms. In Fig. 7, the farms can be separated on the basis of the number of crops that they grow. One group can be said to be ‘rich’ and the other to be ‘poor’. Similarly, two groups are also found for crops, one composed of common crops and the other of rare crops. In Fig. 8, farms are similar and three crops are much more common than the others. Since the difference is quite clear and there are three crops, the ICL criterion for the LBM argues for recognition of a block with only three crops. However, if there are only one or two outlier(s) or if the difference is less clear, this criterion may not separate this (these) outlier(s). This criterion for model selection is not designed for detecting outliers.

### 3.4 Uncovering Outliers Through PCA

#### 3.4.1 Configuration Model

Fix the degree  $(C_i)_{i=1,\dots,n}$  of each farm and  $(F_j)_{j=1,\dots,m}$  of all crops in  $\mathbf{X}$ . The (bipartite) configuration model with parameters  $(C_i)$  and  $(F_j)$  is the uniform distribution over all incidence matrices that leave the degrees  $C_i$  and  $F_j$  unchanged. In the ecological literature, this model is sometimes referred to as the fixed–fixed null model (Connor and Simberloff, 1979; Ulrich and

Gotelli, 2012; Zaman and Simberloff, 2002). In contrast to the LBM, the configuration model takes as a given that some farms might grow many more crops than others and that some crops are more common than others, but apart from that the incidence matrix is sampled uniformly.

In order to simulate according to the configuration model, we use the `tswap` sequential algorithm (Miklós and Podani, 2004) implemented in the `permatswap` function of the R package `vegan`. The practitioner has to select `burnin` and `thinning` parameters large enough that the algorithm explores well the space of incidence matrices. Although the mixing time of the `tswap` algorithm is unknown, the mixing properties of the sequence can be visually checked using the `plot` method of `permatswap`.

### 3.4.2 PCA on Residuals

The expected incidence matrix under the configuration model with degrees  $(C_i)$  and  $(F_j)$  is denoted  $\mathbb{E}_0[\mathbf{X} | (C_i, F_j)]$ . Alternatively,  $\mathbb{E}_0[\mathbf{X} | (C_i, F_j)]$  can be seen as the average overall permutations on the entries of  $\mathbf{X}$  that keeps the degree sequences for both crops and farms unchanged. Then, the residual matrix  $\mathbf{R}$  under the configuration model is the difference between the observed incidence matrix and its expectation under the configuration model

$$R_{ij} = X_{ij} - \mathbb{E}_0[X_{ij} | (C_i, F_j)] \quad (4)$$

If the incidence matrix  $\mathbf{X}$  was drawn according to the configuration model, then  $\mathbf{R}$  would have no particular structure. In order to check the absence of structure, we apply a (non-standardized) PCA on  $\mathbf{R}$ . As it is customary for a PCA, the projection of the rows (i.e. the farms) along the first principal directions allows (i) discovery of groups of farms that effectively cultivate the same types of crops and (ii) detection of outlier farms whose field crop composition is unusual when the effect of farm richness has been removed. As an example, a farm where a very high diversity is cultivated would not necessarily be an outlier, but this farm will be considered as an outlier if it does not grow some very common crops. The projection of the columns of  $\mathbf{R}$  along the first principal directions provides information on outlier crop or groups of crops.

### 3.4.3 Goodness-of-Fit Test of the Configuration Model

Assessing the statistical significance of the PCA is equivalent to testing whether the network  $\mathbf{X}$  has been drawn according to the configuration model. The test rejects the null hypothesis when the largest eigenvalue in



the scree plot is unusually large. More precisely, the test is calibrated by permutations  $\mathbf{X}^P$  of  $\mathbf{X}$  that leaves the degree of each row and column invariant. Denote  $\lambda_{\max}$  the largest singular value of  $\mathbf{R}$  (i.e. the square-root of the largest eigenvalue of  $\mathbf{R}'\mathbf{R}$ ), then the  $p$ -values are obtained by comparing the singular value  $\lambda_{\max}$  to the largest singular values of matrix  $\mathbf{R}^P$  arising from permutations  $\mathbf{X}^P$ .

Under the null hypothesis, the matrix  $\mathbf{R}$  is pure noise and all the singular values of  $\mathbf{R}$  should be small. Under the presence of outliers or of a few groups of farms that preferentially cultivate some crops, the matrix  $\mathbf{R}$  is expected to be the sum of a noisy component and a low-rank component measuring the deviance from the configuration model. As a consequence, the singular value of  $\mathbf{R}$  should be higher under the alternative than under the null hypothesis.

Although calibrated differently, the largest singular value statistic has been applied in other problems of community detection (Bickel and Sarkar, 2015).

### 3.4.4 A New Representation of the Incidence Matrix

Ordering the farms according to the coordinate of their projection along the first principal direction, we denote  $\sigma_1(i)$  the farm index associated with the  $i$ th smallest coordinate. Similarly,  $\sigma_2(j)$  stands for the reordering of the crops according to their projection on the first direction. These permutations  $(\sigma_1, \sigma_2)$  define a new representation  $\mathbf{Y}$  of the incidence matrix:

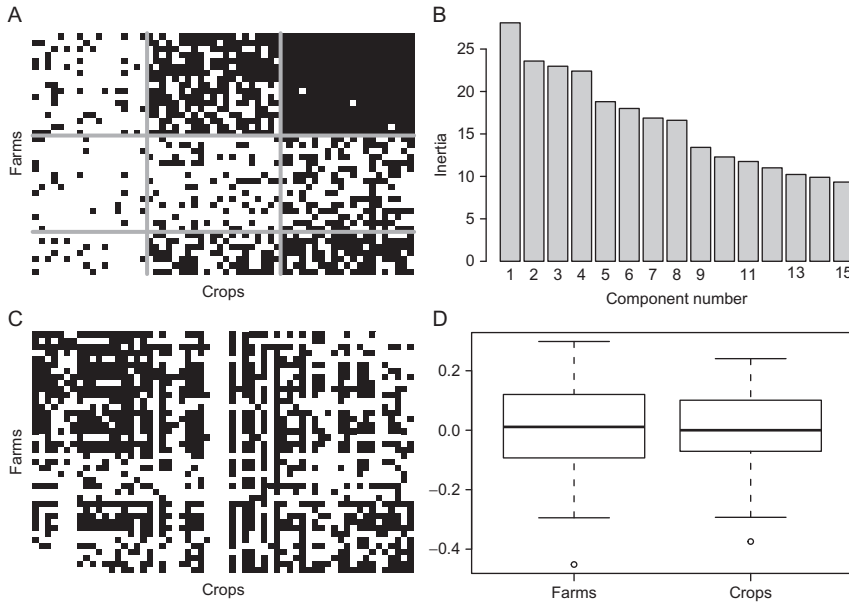
$$Y_{ij} = X_{\sigma_1(i)\sigma_2(j)} \quad (5)$$

This provides a visualization of the incidence matrix alternative to that offered by the LBM approach.

### 3.4.5 Toy Examples

Let us describe three typical examples to understand the behaviour of the above statistics. In all these examples, the number  $n$  of farms is set to 40 and the number  $m$  of crops set to 60.

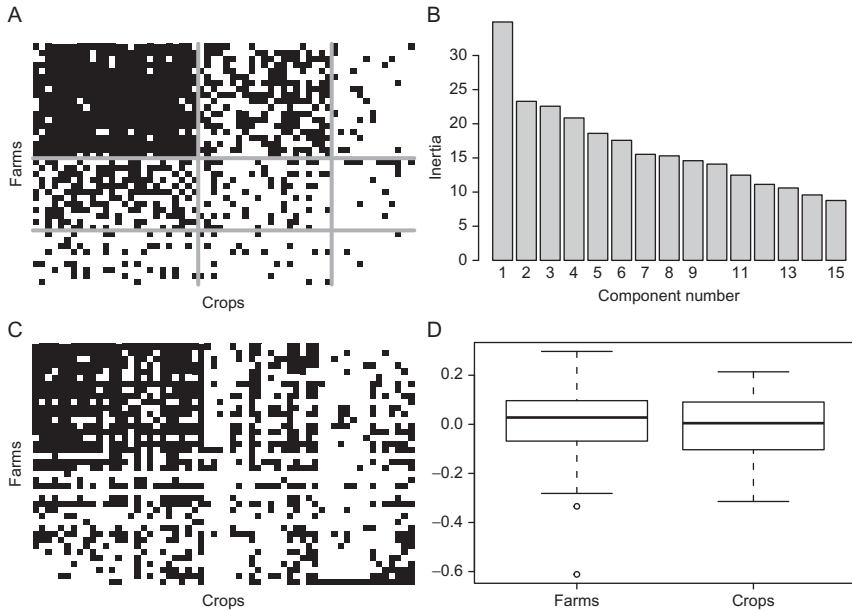
First, we consider a model with degree heterogeneity. For each farm  $i = 1, \dots, n$  and each crop  $j = 1, \dots, m$ , we draw *iid* uniform random variable  $a_i$  and  $b_i$  in  $(0, 1)$ . Then, each entry  $X_{ij}$  is drawn according to a Bernoulli distribution with parameter  $\min(2a_i b_j, 1)$ . As a consequence, the incidence matrix  $\mathbf{X}$  exhibits large degree heterogeneity among farms (resp. crops) with a low  $a_i$  (resp.  $b_i$ ) value and farms (resp. crops) with a high  $a_i$  (resp.  $b_i$ ). It is therefore not unexpected that the LBM estimation procedure (Fig. 9A)



**Figure 9** First example to illustrate the method for uncovering outliers through principal component analysis: degree heterogeneity. (A) The LBM representation, (B) the scree plot of the PCA residuals, (C) the representation of the incidence matrix according to the PCA ordering (Eq. 5) and (D) the boxplots of the PCA first coordinates.

recovers several groups of crops and farms. The  $p$ -value of configuration model from Section 3.4.3 equals 0.54. Again, this is not surprising, since this incidence matrix has been sampled from a model similar to the configuration model. This implies that the block structure found by the LBM method can be explained by the degree heterogeneity. As the configuration model residuals are completely random here, both the PCA scree plot (Fig. 9B) and the representation (Eq. 5) of the incidence matrix (Fig. 9C) are uninformative. No farms and no crops have outlier PCA coordinates (Fig. 9D).

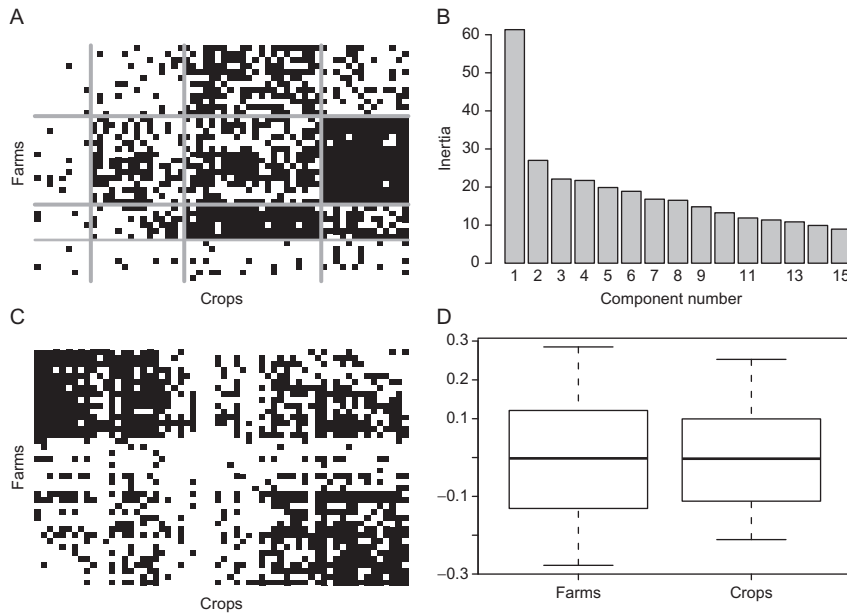
In the second example, we draw the incidence matrix  $\mathbf{X}$  as above. Then, we replace each entry of the first row by independent Bernoulli random variables with parameter 0.5. As a consequence, the first farm is assumed to have a completely different behaviour from all the other farms, as it grows crops regardless of their scarcity ( $b_j$ ) in the village. The LBM representation (Fig. 10A) is close to that of the first example (Fig. 9A). The  $p$ -value of the configuration test is smaller than  $10^{-3}$ , and the scree plot exhibits an unusually large first eigenvalue (Fig. 10B). The first farm is therefore detected as an outlier by the first coordinate representation (Fig. 10D).



**Figure 10** Second example to illustrate the method for uncovering outliers through principal component analysis: outlier. (A) The LBM representation, (B) the scree plot of the PCA residuals, (C) the representation of the incidence matrix according to the PCA ordering (Eq. 5) and (D) the boxplots of the PCA first coordinates.

Finally, the PCA-based representation (Fig. 10C) highlights the unusual behaviour of this farm.

In the last example, we draw random variables  $a_i$  and  $b_j$  as above. Then, the farms are divided into two groups  $A_1$  and  $A_2$  of size  $n/2$  and the crops are divided into two groups  $B_1$  and  $B_2$  of size  $m/2$ . Then, the entry  $X_{ij}$  is drawn according to Bernoulli distribution with parameter  $\min(p_{\text{in}}2a_ib_j, 1)$  if  $(i, j) \in A_1 \times B_1$  or  $(i, j) \in A_2 \times B_2$  and parameter  $\min(p_{\text{out}}2a_ib_j, 1)$  if  $(i, j) \in A_1 \times B_2$  or  $(i, j) \in A_2 \times B_1$  with  $p_{\text{in}} = 1.4$  and  $p_{\text{out}} = 0.6$ . Intuitively, the farms from  $A_1$  (resp.  $A_2$ ) preferentially grow crops from  $B_1$  (resp.  $B_2$ ), but the model also allows the degree of the farm and of each crop to be heterogeneous inside the blocks. As a consequence, this model, called degree-corrected, is neither a LBM with  $2 \times 2$  blocks nor a configuration model but a blend of them. The LBM estimation method recovers too many blocks (Fig. 11A) by grouping farms or crops that are in the same group and have similar degrees. The  $p$ -value for the configuration test is found to be smaller than  $10^{-3}$ . This is corroborated by the fact that the scree plot exhibits an unusually large first eigenvalue (Fig. 11B). Contrary to the previous example, this unusually



**Figure 11** Third example to illustrate the method for uncovering outliers through principal component analysis: blocks. (A) The LBM representation, (B) the scree plot of the PCA residuals, (C) the representation of the incidence matrix according to the PCA ordering (Eq. 5) and (D) the boxplots of the PCA first coordinates.

large singular value is not due to outliers (Fig. 11D) but to the presence of a block structure. The PCA-based matrix representation highlights the presence of these two groups of farms and crops (Fig. 11C).

### 3.5 Measuring Originality of Farms' Contributions Through Diversity Measures

We will now focus our attention on the distribution of cultivated crop diversity at the level of the sampled location (the village). As mentioned in previous sections, some farms may grow many more crops than others (hence, the high variance in degree among farms in the bipartite network). A question that remains unanswered is whether low-degree farms contribute effectively more or less than high-degree farms to the overall cultivated diversity—‘effectively more’ being understood as contributing more than expected if crops were chosen randomly from the pool of crops cultivated in the village. In other words, the question is now whether low-degree farms cultivate the most frequent crops in the village only (common crops) or

contribute disproportionately to crop diversity by focusing only on crops that are cultivated on very few farms (rare crops).

### 3.5.1 Theoretical Framework

Further expanding the notations introduced in [Section 3.1](#), we denote  $p_{ij}$  the weight associated with the interaction between farm  $i$  and crop  $j$  among all interactions of farm  $i$ :

$$p_{ij} = \frac{X_{ij}}{C_i} \quad (6)$$

The proportion of all the connections in the network that are due to farm  $i$  or crop  $j$  are, respectively, noted  $q_i$  and  $h_j$ :

$$q_i = \frac{C_i}{N} \quad (7)$$

$$h_j = \frac{F_j}{N} \quad (8)$$

We note  $H_i$  the diversity of crops cultivated on farm  $i$ , as measured by Shannon entropy:

$$H_i = - \sum_j p_{ij} \log p_{ij} = \log C_i \quad (9)$$

The average diversity among farms, weighted by the importance  $q_i$  of each farm, is denoted  $H_\alpha$ :

$$H_\alpha = \sum_i q_i H_i = \frac{1}{N} \sum_i C_i \log C_i \quad (10)$$

The diversity of crops cultivated by all farms, when taken together and weighted by the importance  $q_i$  of each farm, is noted  $H_T$  and reads as:

$$\begin{aligned} H_T &= - \sum_j \left[ \sum_i q_i p_{ij} \right] \log \left[ \sum_i q_i p_{ij} \right] = - \sum_j h_j \log h_j \\ &= \log N - \frac{1}{N} \sum_j F_j \log F_j \end{aligned} \quad (11)$$

The difference between  $H_T$  and  $H_\alpha$  is the turnover in diversity among farms or  $\beta$  diversity, noted  $H_\beta$ :

$$H_\beta = H_T - H_\alpha = \log N - \frac{1}{N} \sum_j F_j \log F_j - \frac{1}{N} \sum_i C_i \log C_i \quad (12)$$

$H_\beta$  can be further decomposed into individual turnover components,  $H_{iT}$ :

$$H_\beta = \sum_i q_i H_{iT} \quad (13)$$

where  $H_{iT}$  measures the ‘originality’ of farm  $i$  portfolio of crops when compared to the overall diversity of cultivated crops. An expression for  $H_{iT}$  can be found (Lande, 1996):

$$H_{iT} = - \sum_j p_{ij} \log \frac{C_i F_j}{N} \quad (14)$$

### 3.5.2 Measuring the Diversity Cultivated by Crop-Poor and Crop-Rich Farms

We now focus on measuring the evenness of crops cultivated by a subset  $I$  of farms. More specifically, because we are interested in the subset of the most crop-poor or crop-rich farms, we will assume that the set  $I$  contains all farms belonging to a certain quantile of the distribution of  $S_i$ . The evenness of crops cultivated on farms in set  $I$  is noted  $E_I$  and reads as

$$E_I = - \frac{\sum_j \left[ \sum_{i \in I} q_{i,I} p_{ij} \right] \log \left[ \sum_{i \in I} q_{i,I} p_{ij} \right]}{\log(m)}; \quad q_{i,I} = \frac{C_i}{\sum_{i \in I} C_i}. \quad (15)$$

The evenness  $E_I$  is the diversity of crops cultivated on all farms in set  $I$  divided by the logarithm of the total number  $m$  of crops cultivated in the village. It measures the equity of the distribution of crops cultivated on farms in  $I$ .

In order to assess whether the evenness is greater in crop-rich farms than crop-poor farms, we compare the value of  $E_{\text{Rich}} - E_{\text{Poor}}$  to that of all realizations of the incidence matrix  $\mathbf{X}$  under the configuration model (i.e. randomizing connections given degree sequences for both crops and farms) by a permutation test.

### 3.5.3 Measuring the Impact of Crop-Poor and Crop-Rich Farms

We now focus on measuring the  $\beta$  diversity  $H_{\beta,I}$  due to the contribution of a subset  $I$  of farms. As previously, the subset  $I$  is made up of the most

crop-poor or crop-rich farms. We can give an explicit formula for  $H_{\beta,I}$  (Lande, 1996):

$$H_{\beta,I} = \sum_{i \in I} q_i H_{iT} = - \sum_i q_i \log q_i + \frac{1}{N} \sum_j \left[ \sum_{i \in I} X_{ij} \right] \log \left( \frac{1}{F_j} \right) \quad (16)$$

The first term in the right-hand side of Eq. (16) relies on the expression of the  $\alpha$  diversity  $H_{\alpha,I}$  due to farms in subset  $I$ :

$$H_{\alpha,I} = \frac{1}{N} \sum_{i \in I} C_i \log C_i = \frac{\sigma_I \log N}{N} + \sum_{i \in I} q_i \log q_i \quad (17)$$

where  $\sigma_I$  is the ‘volume’ of interactions due to farms belonging to subset  $I$ :

$$\sigma_I = \sum_{i \in I} C_i \quad (18)$$

The second term depends on the correlation between a crop degree  $F_j$  and the number of farms within the set  $I$  who possess this crop, noted  $\varphi_{j,I}$ :

$$\varphi_{j,I} = \sum_{i \in I} X_{ij} \quad (19)$$

Plugging Eqs. (17)–(19) into Eq. (16) yields the following expression for  $H_{\beta,I}$ :

$$H_{\beta,I} = \frac{\sigma_I \log N}{N} - H_{\alpha,I} - \frac{1}{N} \sum_j \varphi_{j,I} \log F_j \quad (20)$$

The quantity  $D_I = \frac{1}{N} \sum_j \varphi_{j,I} \log F_j$  measures the deficit of originality displayed by the farms in subset  $I$  that is due to their cultivation of ‘common crops’.

Again, we assess the significance of  $H_{\beta,I}$  by a permutation test based on the configuration model. As the set  $I$  contains all farms belonging to a certain quantile of the distribution of  $C_i$ , all realizations of the incidence matrix  $\mathbf{X}$  under the configuration model preserve the set of  $C_i$  values to be found in  $I$ .

As a consequence, the quantity  $\frac{\sigma_I \log N}{N} - H_{\alpha,I}$  in the right-hand side of Eq. (16) is invariant with respect to the configuration model. The quantity  $D_I$  in the right-hand side of Eq. (16), however, does not satisfy this invariance. Thus, values of  $H_{\beta,I}$  that are unusually large for the configuration model mean that farms in subset  $I$  contribute more to cultivated biodiversity than expected by the number of types cultivated on farms in  $I$ .

### 3.5.4 Measuring Originality of Farms' Contributions Through Diversity Measures on Toy Examples

#### 3.5.4.1 Simulation Model

Two groups of farms are considered: crop-rich (40% of farms) and crop-poor (60% of farms). The crops are divided into two groups with same size: rare and common, consistently with the definition provided at the beginning of Section 3.5. The entries of the incidence matrix are generated as *iid.* Bernoulli random variables with probability  $p_{ij}$  (corresponding to farm  $i$  and crop  $j$ ) given by:

$$\text{logit}(p_{ij}) = \mu + \alpha(L_i) + \beta(K_j) + \gamma(L_i: K_j)$$

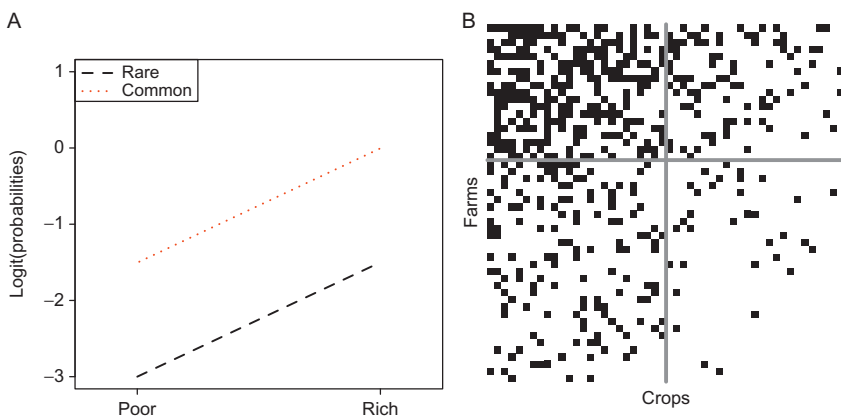
where  $\text{logit}$  is the function  $x \mapsto \log(x/(1-x))$ ,  $L_i$  indicates the group of farm  $i$ ,  $K_j$  the group of crop  $j$  and parameters  $\mu$ ,  $\alpha$ s,  $\beta$ s,  $\gamma$ s are:

$$\alpha(\text{poor}) = \beta(\text{rare}) = \gamma(\text{poor, rare}) = \gamma(\text{rich, rare}) = \gamma(\text{poor, common}) = 0$$

to ensure identifiability. The interaction term  $\gamma(\text{rich, common})$  then drives the respective contributions to diversity of crop-rich and crop-poor farms. Indeed, if the value of this term is zero, the effect of being crop-rich for growing a rare or a common variety will be the same.

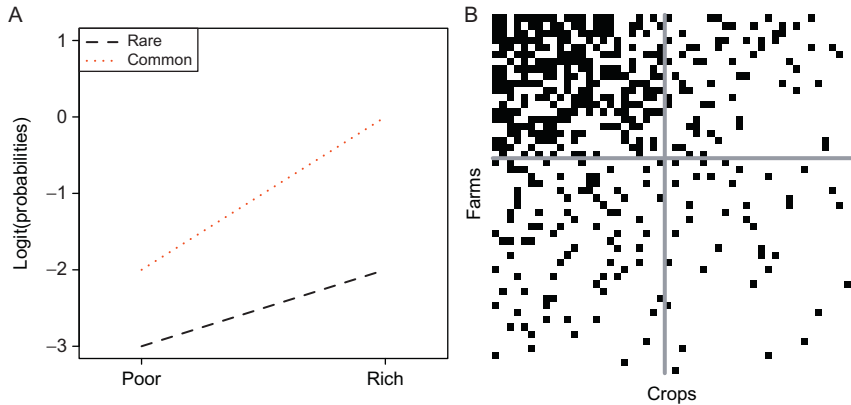
#### 3.5.4.2 Three Contrasted Toy Examples

Figures 12–14 correspond, respectively, to the three following settings:

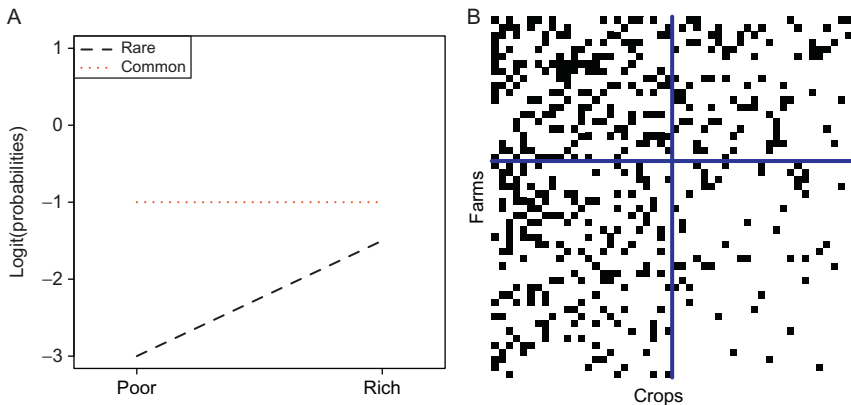


**Figure 12** Toy example with equal contributions to diversity of crop-rich and crop-poor farms.  $\mu = -3$ ,  $\alpha(\text{rich}) = \beta(\text{common}) = 1.5$ ,  $\gamma(\text{rich, common}) = 0$ . (A) Probabilities that a crop is grown on a farm and (B) incidence matrix.





**Figure 13** Toy example with greater contribution to diversity of crop-poor farms.  $\mu = -3$ ,  $\alpha(\text{rich}) = \beta(\text{common}) = (\text{rich}, \text{common}) = 1$ . (A) Probabilities that a crop is grown on a farm and (B) incidence matrix.



**Figure 14** Toy example with greater contribution to diversity of crop-rich farms.  $\mu = -3$ ,  $\alpha(\text{rich}) = -\gamma(\text{rich}, \text{common}) = 1.5$ ,  $\beta(\text{common}) = 2$ . (A) Probabilities that a crop is grown on a farm and (B) incidence matrix.

1. Parameters fixed to  $\mu = -3$ ,  $\alpha(\text{rich}) = \beta(\text{common}) = 1.5$ ,  $\gamma(\text{rich}, \text{common}) = 0$ . The crop-rich and crop-poor farms have the same contribution to diversity with respect to their own crop richness. This is ensured by setting the interaction term  $\gamma(\text{rich}, \text{common})$  to 0.
2. Parameters fixed to  $\mu = -3$ ,  $\alpha(\text{rich}) = \beta(\text{common}) = \gamma(\text{rich}, \text{common}) = 1$ . The crop-poor farms have a greater contribution to diversity, since they grow rare crops and common crop with nearly the same probability, whereas for crop-rich farms, the probability of growing rare crops is clearly smaller than the probability of growing common crops.

3. Parameters fixed to  $\mu = -3$ ,  $\alpha(\text{rich}) = -\gamma(\text{rich, common}) = 1.5$ ,  $\beta(\text{common}) = 2$ . The crop-rich farms have a greater contribution to diversity, as they exhibit the same ability of growing rare and common crops. As shown in Figs. 12–14, these three settings are consistent with what a crop-poor or a crop-rich farm and a rare or a common crop are expected to be.

Results in Table 4 are coherent with the intuitive expectations based on these three models. For the first case, nothing was found to contribute significantly to the distribution of crop diversity. For the two other cases, the tests on evenness and on the contributions to diversity of crop-rich and crop-poor farms agreed. Indeed, in the case of Fig. 13, for instance, the crop-rich farms are found to contribute less than expected to diversity (null hypothesis rejected on the left side), the crop-poor farms are found to contribute more than expected to diversity (null hypothesis rejected on the right side) and the difference of evenness is found to be significantly smaller than expected (null hypothesis rejected on the left side). Based on these three toy examples, a power study was conducted and its results are presented in Appendix.



## 4. PATTERNS OF LOCAL CROP DIVERSITY: RESULTS OF THE META-ANALYSIS

All the aforementioned methods have been applied to the 50 datasets of the meta-analysis. The results are summarized in Tables 5 and 6 for species and infra-species diversity, respectively.

### 4.1 Variability of Farms' and Crops' Degrees

The aim of this section is to detect over-dispersion (significant test on the right) or under-dispersion (significant test on the left) of degree distribution for farms and crops, respectively, following the methodology introduced in Section 3.2. Two null hypotheses ( $H_0$ ) are tested:

1. diversity at species and infra-species level is randomly distributed (i.e. according to a binomial distribution) among farms from the same village (homogeneity of the farm degrees) and
2. crop richness is randomly distributed within the same village (homogeneity of the crop degrees).

#### 4.1.1 Species Diversity

For farms,  $H_0$  was rejected on the right side (16 times over the 18 tested datasets) for the variability of farms' degree (Table 5). There was only

**Table 4**  $p$ -Values for the Contribution Tests Applied to the Three Examples Presented in Figs. 12–14

	Evenness Difference Rich–Poor Farms	Significance		Crop–Rich Farms Contribution	Significance		Crop–Poor Farms Contribution	Significance	
		Left	Right		Left	Right		Left	Right
Fig. 12	0.036	0.433	0.567	0.628	0.587	0.413	0.643	0.413	0.587
Fig. 13	−0.003	<0.001	>0.999	0.638	<0.001	>0.999	0.64	>0.999	<0.001
Fig. 14	0.036	0.989	0.011	0.715	0.996	0.004	0.809	0.004	0.996

**Table 5** Statistical Results Obtained for the 18 Datasets Dealing with Specific Diversity

Dataset	Variance of Farms' Degree	Significance		Variance of Species' Degree	Significance		LBM Farm Cluster Number	LBM Species Cluster Number	Normalized First Singular Value	Significance	Evenness Difference Rich-Poor Farms	Significance		Crop-Rich Farms Contribution	Significance		Crop-Poor Farms Contribution	Significance	
		Left	Right		Left	Right						Left	Right		Left	Right			
CL-M01	0.579	<0.001	NS	74.138	NS	<0.001	1	3	0.49	NS	0.167	NS	<0.001	2.187	NS	<0.001	1.796	NS	NS
CV-M01	1.843	NS	<0.05	8.068	NS	<0.001	2	2	2.21	<0.05	0.077	NS	NS	3.054	NS	NS	2.824	NS	NS
EG-M05	7.708	NS	<0.001	6.023	NS	<0.001	2	3	1.07	NS	0.061	NS	NS	3.751	NS	NS	3.538	NS	<0.05
EG-M08	10.207	NS	<0.001	6.3	NS	<0.001	3	3	0.98	NS	0.078	<0.01	NS	3.887	<0.05	NS	3.606	NS	<0.001
OC-M02	6.471	NS	<0.001	4.707	NS	<0.001	2	2	4.7	<0.001	0.08	NS	NS	4.429	NS	NS	4.147	NS	NS
OC-M04	8.353	NS	<0.001	10.016	NS	<0.001	3	3	4.05	<0.05	0.127	NS	<0.001	4.361	NS	<0.001	3.968	<0.001	NS
OC-M05	5.346	NS	<0.001	6.285	NS	<0.001	2	2	2.84	<0.01	0.094	NS	NS	4.323	NS	NS	4.059	NS	NS
OC-M06	5.546	NS	<0.001	5.205	NS	<0.001	2	2	0.75	NS	0.102	NS	NS	4.194	NS	NS	3.859	NS	NS
OC-M07	8.431	NS	<0.001	4.31	NS	<0.001	3	2	2.84	<0.01	0.174	NS	<0.001	4.142	NS	<0.001	3.668	<0.001	NS
OC-M09	16.755	NS	<0.001	2.859	NS	<0.001	2	2	0.17	NS	0.19	NS	NS	4.131	NS	NS	3.734	NS	NS
OC-M10	13.601	NS	<0.001	4.007	NS	<0.001	2	2	1.03	NS	0.144	NS	NS	4.157	NS	NS	3.826	NS	NS
OC-M11	5.801	NS	<0.001	18.977	NS	<0.001	2	4	3.54	<0.01	0.085	NS	<0.001	4.583	NS	<0.001	4.202	<0.001	NS
OC-M12	5.187	NS	<0.001	4.529	NS	<0.001	2	2	1.76	<0.05	0.077	NS	<0.001	4.627	NS	<0.001	4.31	<0.001	NS
OC-M13	10.607	NS	<0.001	9.764	NS	<0.001	3	3	2.28	<0.05	0.093	NS	NS	4.602	NS	NS	4.246	<0.05	NS
OC-M14	9.702	NS	<0.001	6.367	NS	<0.001	3	3	2.56	<0.01	0.081	NS	NS	4.587	NS	<0.05	4.266	NS	NS
OC-M16	9.14	NS	<0.001	6.513	NS	<0.001	3	2	1.61	NS	0.11	NS	NS	4.623	NS	NS	4.219	NS	NS
SC-M05	1.671	NS	NS	9.74	NS	<0.001	1	3	0.35	NS	0.11	NS	NS	3.468	NS	<0.001	3.112	NS	NS
SC-M06	4.595	NS	<0.001	3.662	NS	<0.001	2	2	1.1	NS	0.066	NS	NS	2.678	NS	NS	2.471	NS	NS

NS: non-significant, <0.05: *p*-values ranging from 0.01 to 0.05, <0.01: *p*-values ranging from 0.001 to 0.01 and <0.001: *p*-values lower than 0.001.

**Table 6** Statistical Results Obtained for the 33 Datasets Dealing with Infra-Specific Diversity

Dataset	Variance of Farms' Degree	Significance		Variance of Landraces' Degree	Significance		LBM Farm Cluster Number	LBM Landraces Cluster Number	Normalized First Singular Value	Significance	Evenness Difference Rich-Poor Farms	Significance		Crop-Rich Farms' Contribution	Significance		Crop-Poor Farms' Contribution	Significance	
		Left	Right		Left	Right						Left	Right		Left	Right			
AB-M02	1.485	NS	NS	3.456	NS	<0.001	1	2	0.82	NS	0.116	NS	NS	2.951	NS	NS	2.667	NS	NS
CL-M02	0.338	<0.001	NS	33.193	NS	<0.001	1	3	2.05	<0.05	0.057	NS	NS	1.978	NS	NS	1.913	NS	NS
CV-M02	0.963	NS	NS	5.731	NS	<0.001	1	2	0.76	NS	0.138	NS	NS	2.943	NS	NS	2.545	<0.05	NS
DJ-M003a	1.191	NS	NS	14.849	NS	<0.001	1	3	0.5	NS	0.224	NS	<0.05	0.449	NS	NS	0.371	NS	NS
DJ-M003b	1.207	NS	NS	17.837	NS	<0.001	1	3	0.69	NS	0.142	NS	NS	0.649	NS	NS	0.561	NS	NS
DJ-M003c	1.164	NS	NS	8.074	NS	<0.001	1	2	0.03	NS	0.194	NS	NS	0.564	NS	NS	0.479	NS	NS
DJ-M003d	0.751	NS	NS	10.915	NS	<0.001	1	2	0.43	NS	0.216	NS	NS	0.497	NS	NS	0.424	NS	NS
DJ-M009a	0.995	NS	NS	13.273	NS	<0.001	1	2	0.68	NS	0.163	NS	NS	0.385	NS	NS	0.463	NS	NS
DJ-M009b	1.033	NS	NS	9.983	NS	<0.001	1	2	0.66	NS	0.378	NS	NS	0.585	NS	NS	0.39	NS	NS
DJ-M009c	0.966	NS	NS	15.084	NS	<0.001	1	2	1.6	NS	0.422	NS	NS	0.648	NS	NS	0.39	NS	NS
DJ-M009d	0.975	NS	NS	5.285	NS	<0.001	1	2	0.57	NS	0.318	NS	NS	0.337	NS	NS	0.362	NS	NS
DJ-M009e	0.848	NS	NS	13.349	NS	<0.001	1	2	0.33	NS	0.271	NS	NS	0.468	NS	NS	0.389	NS	NS
DJ-M009f	1.349	NS	NS	10.697	NS	<0.001	1	2	0.07	NS	0.313	NS	NS	0.596	NS	NS	0.421	NS	NS
DJ-M012a	0.656	<0.05	NS	16.079	NS	<0.001	1	3	1.62	NS	0.124	NS	<0.001	3.206	NS	<0.001	2.794	NS	NS
DJ-M012b	1.18	NS	NS	6.892	NS	<0.001	1	2	0.07	NS	0.086	NS	<0.05	3.576	NS	NS	3.468	NS	NS
DJ-M015a	0.304	<0.001	NS	88.723	NS	<0.001	1	3	0.81	NS	0.097	NS	NS	2.035	NS	NS	1.798	NS	<0.05
DJ-M015b	0.542	<0.05	NS	9.976	NS	<0.001	1	2	1.82	NS	0.099	NS	NS	1.979	NS	NS	1.744	NS	NS
DJ-M018a	0.367	<0.001	NS	9.875	NS	<0.001	1	2	3.95	<0.05	0.246	NS	<0.001	2.354	NS	<0.001	1.782	<0.001	NS
DJ-M018b	0.297	<0.001	NS	45.012	NS	<0.001	1	3	3.82	<0.001	0.153	NS	NS	1.883	NS	NS	1.519	NS	NS
DJ-M030	0.44	<0.001	NS	48.912	NS	<0.001	1	3	3.92	<0.05	0.233	NS	<0.001	1.62	NS	NS	1.21	<0.01	NS

*Continued*

**Table 6** Statistical Results Obtained for the 33 Datasets Dealing with Infra-Specific Diversity—cont'd

Dataset	Variance of Farms' Degree	Significance		Variance of Landraces' Degree	Significance		LBM Farm Cluster Number	LBM Landraces Cluster Number	Normalized First Singular Value	Significance	Evenness Difference Rich-Poor Farms	Significance		Crop-Rich Farms' Contribution	Significance		Crop-Poor Farms' Contribution	Significance	
		Left	Right		Left	Right						Left	Right		Left	Right			
DJ-M036	0.637	<0.05	NS	13.24	NS	<0.001	1	2	1.45	NS	0.198	NS	<0.001	2.328	NS	<0.05	1.84	<0.001	NS
DJ-M039a	0.928	NS	NS	0.94	NS	NS	1	1	0.25	NS	0.234	NS	NS	2.236	NS	NS	1.975	NS	NS
DJ-M045b	0.205	<0.001	NS	0.736	NS	NS	1	1	0.03	NS	0.08	NS	NS	2.384	NS	NS	2.321	NS	NS
DJ-M045c	0.746	NS	NS	0.491	NS	NS	1	1	0.45	NS	0.211	NS	NS	2.377	NS	NS	2.221	NS	NS
DJ-M045d	0.357	<0.05	NS	2.295	NS	<0.05	1	1	0.8	NS	0.356	NS	NS	2.099	NS	NS	1.679	NS	NS
JW-M07	0.853	NS	NS	10.789	NS	<0.001	1	3	0.51	NS	0.068	NS	NS	2.644	<0.05	NS	2.465	NS	NS
JW-M08	0.536	<0.01	NS	15.333	NS	<0.001	1	4	0.74	NS	0.115	NS	NS	2.663	NS	NS	2.334	<0.05	NS
JW-M09	0.816	NS	NS	15.148	NS	<0.001	1	4	0.92	NS	0.043	NS	NS	2.873	NS	NS	2.736	NS	NS
JW-M10	0.949	NS	NS	5.73	NS	<0.001	1	2	1.68	NS	0.053	NS	NS	2.76	NS	NS	2.587	NS	NS
ME-M01	3.9	NS	<0.001	7.395	NS	<0.001	2	3	0.84	NS	0.112	NS	NS	3.997	NS	NS	3.649	NS	NS
SC-M04	8.172	NS	<0.001	3.842	NS	<0.001	2	2	0.45	NS	0.099	NS	NS	3.454	NS	NS	3.216	NS	NS
SC-M07	3.335	NS	<0.001	2.998	NS	<0.001	2	2	0.6	NS	0.126	NS	NS	3.188	NS	NS	3.021	NS	<0.001

NS: non-significant, <0.05: *p*-values ranging from 0.01 to 0.05, <0.01: *p*-values ranging from 0.001 to 0.01 and <0.001: *p*-values lower than 0.001.

one case where the test was not significant on both sides (SC-M05) and one case where the test was rejected on the left side (CL-M01). These results indicate that the number of species grown per farm from the same village is generally over-dispersed, with few farms growing more species than expected. For the variability in degree of species, this pattern was even stronger, with a systematically over-dispersed degree distribution.

#### **4.1.2 *Infra-Species Diversity***

For farms at the infra-specific level, the pattern is completely different as  $H_0$  is rejected on the right side only 3 times over the 32 tested datasets (ME-M01, SC-M04 and SC-M07), and 11 times on the left side (Table 4). These results indicate an under-dispersion of the degree distribution when we consider the distribution of landraces at the village scale. For degree of landraces,  $H_0$  is mostly rejected on the right side with 29 times over the 32 datasets, indicating, as for the species level, an over-dispersion of the degree distribution.

## **4.2 Structure Detection Through Model-Based Clustering (LBM)**

In this section, we seek to detect the existence of patterns within inventory datasets at the village scale using LBM as explained in Section 3.3.

#### **4.2.1 *Species Diversity***

The clustering method applied to the different datasets detected from one to three clusters for the farms and from two to three clusters for the species (Table 5 and Fig. A1). These results are similar to the toy example illustrated in Fig. 7. Therefore, the clustering seems mostly driven by the heterogeneity in degree of both farms and species. Farms were clustered together because they grow almost the same species. In the case of two clusters for farms, we then define the ‘crop-poor’ farm cluster as the one with the lower density and the ‘crop-rich’ farm cluster as the one with the higher density. In the case of two groups for the species, we define the ‘rare species’ cluster as the one with the lower number of links and the frequent species cluster as the one with the higher number of links.

#### **4.2.2 *Infra-Species Diversity***

The clustering method detected from one to two clusters for the farms and from one to four clusters for landraces (Table 6 and Figs. A2 and A3). For four datasets (DJ-M039a, DJ-M045b, DJ-M045c and DJ-M045d), only one

cluster was detected both for farms and landraces (Table 4). These results of low clustering are consistent with the low variability of the degrees both for the farms and the landraces observed in Section 4.1. Similarly, 26 additional datasets with under-dispersion had only one block for the farms. These findings indicate that for landrace diversity, a lower heterogeneity is generally observed among farms where nearly the same landraces are grown. Only three datasets showed two blocks for the farms (ME-M01, SC-M04 and SC-M07). Nevertheless, it is still possible to distinguish between frequent and rarer landraces.

### 4.3 Outlier Detection Through PCA

We then applied a PCA to detect farms that are ‘outliers’ in terms of species and infra-species diversity. See Section 3.4 for methodological details.

#### 4.3.1 Species Diversity

Using the test introduced in Section 3.4.3,  $H_0$  was rejected 9 times over the 18 datasets at  $\alpha = 0.05$ , highlighting the existence of outliers. These outliers are generally two or three farms per dataset (Fig. A4), which can be characterized as farms where a different subset of species is grown compared to other farms with an equivalent degree, belonging to the same cluster.

#### 4.3.2 Infra-Species Diversity

$H_0$  was rejected for 4 datasets over the 32 datasets (CL-M02, DJ-M018a, DJ-M018b and DJ-M030, Fig. A5). These results indicate that in addition to growing almost the same number of landraces, the same portfolio of landraces is grown globally by all farms from the same village. Note that for these four datasets, only one cluster was detected with the LBM (CL-M02, DJ-M018a, DJ-M018b and DJ-M030). Therefore, in this case, we have farms with a particular subset of landraces and having an equivalent degree.

### 4.4 Farms’ Contributions to Local Diversity

In the analyses reported in this section, farms were separated into ‘crop-rich’ farms and ‘crop-poor’ farms according to their degree in such a way that arbitrarily 40% of farms were classified as ‘crop-rich’. The method described below is not highly sensitive to this threshold value, except for extreme values.

Evenness ( $E$ ) and contribution ( $H_\beta$ ) were computed for each of these two groups as explained in Section 3.5.



#### 4.4.1 Species Diversity

The tests on the difference between  $E_{\text{rich}}$  and  $E_{\text{poor}}$  revealed that crop-rich farms had a significantly higher evenness in five cases (CL-M01, OC-M04, OC-M07, OC-M11 and OC-M12). The group of crop-poor farms contributed significantly more than that of crop-rich farms in only one case (EG-M08).  $H_0$  was not rejected in the other cases, indicating that no significant difference in terms of contribution to the global diversity by the crop-rich group of farms compared to the crop-poor group.

Our findings on the difference between  $E_{\text{rich}}$  and  $E_{\text{poor}}$  converge with the test of the contributions of crop-rich and crop-poor farms. Indeed, in five cases when the first test was significant on the right side (i.e. a significantly higher contribution to the global diversity by the crop-rich farms than the crop-poor farms), we observed that some crop-rich groups did indeed contribute significantly to the global diversity and that some crop-poor groups contributed significantly less than expected in four of the five cases (Table 5). Two additional datasets showed a significant contribution of the crop-rich farms (OC-M14 and SC-M05) and one additional dataset showed that the crop-poor farms contributed significantly less than expected (OC-M13). The crop-poor farms contributed significantly more than expected in only two cases. In one of these cases (EG-M05), the result is consistent with that of the test on evenness. In the other case (EG-M08), crop-poor farms only showed a significant contribution to global diversity and not to evenness (EG-M08).

#### 4.4.2 Infra-Species Diversity

The tests of the difference between  $E_{\text{rich}}$  and  $E_{\text{poor}}$  farms revealed that crop-rich farms had a significantly higher evenness in six cases (Table 6; DJ-M003a, DJ-M012a, DJ-M012b, DJ-M018a, DJ-M030 and DJ-M036).  $H_0$  was not rejected in the other cases, indicating no significant difference in evenness between crop-rich and crop-poor farms. These results were not always convergent with the results of the tests on the contributions of the crop-rich and crop-poor farm groups to diversity at the village level. These latter tests gave convergent results (a significant contribution of a few crop-rich farms to the global diversity) in only two cases (DJ-M018a and DJ-M036) of the six in which the evenness difference was significant. In one additional dataset, few farms from the crop-rich group contributed significantly less than expected (JW-M07). In one additional dataset, the crop-poor farms contributed significantly more than expected (DJ-M012a). In three additional datasets, few farms from the crop-poor

group contributed significantly less than expected (CV-M02, DJ-M030 and JW-M08). Finally, in two datasets, the crop-poor farms contributed significantly more than expected (DJ-M0015a and SC-M07).



## 5. DISCUSSION

### 5.1 Contrasted Patterns of Local Crop Diversity at the Species and Infra-Species Levels

Applying a set of network-based methods to a meta-dataset of crop diversity reveals distinct sources of heterogeneity in terms of crop distribution at the local scale:

- i. crop diversity among farms is generally more heterogeneous at the specific level than at the infra-specific level;
- ii. heterogeneity in farms' degrees is one explanation for this heterogeneity, with blocks of low-diversity farms and of high-diversity farms (the same pattern is observed for species and landraces, with blocks of common crops and blocks of rarer crops);
- iii. outlier farms with unusual portfolios are another source of heterogeneity and
- iv. both low-diversity or high-diversity farms can contribute disproportionately to local diversity by growing rare varieties.

We suggest two main explanations for these general results: heterogeneity in data collection methods and diversity of socio-ecological and environmental contexts. As datasets were collected following different protocols, differences in sampling effort could have an influence on the observed diversity (Perrault-Archambault and Coomes, 2008). An additional source of heterogeneity exists specially at the infra-specific level in the way in which landraces are named and how they are grouped together when they show strong evidence of being the same biocultural object. Nevertheless, a subset of the datasets for landraces were collected in the context of a coordinated global partnership of researchers in order to use a standardized protocol and the same sampling strategy during data collection (Jarvis et al., 2008), and datasets collected in this context also show different patterns (Table 6: DJ-M012a, DJ-M012b, DJ-M015a, DJ-M015b, DJ-M018a, DJ-M018b, DJ-M030, DJ-M036, DJ-M039a, DJ-M045b, DJ-M045c and DJ-M045d). Consequently, variation in the agro-ecological and the socio-cultural contexts, and interactions therein, is likely to have strongly shaped the distribution of local crop diversity.

More specifically, our findings of over-dispersion of the degrees at the specific level and of under-dispersion at the infra-specific level are strengthened by the results of classification using LBM. Indeed, in the cases of over-dispersion, two or three blocks of farms are detected whereas for cases of under-dispersion, only one block of farms is detected. Convergence of the results between these two approaches indicates that the variability of the degree distribution is probably the main driver of block structure. It thus makes sense to use as null model a configuration model, controlling for degree, because this would allow assessment of whether other structural drivers, in addition to the degree, act to shape the patterns of diversity. From an ethnobiological or agro-ecological point of view, the block detection means that farms can be distinguished according to the level of diversity they grow. We identify high-diversity and low-diversity farms. Similarly, for crops, we identify common species/landrace (present in fields of most farms) and rare species/landraces (grown on few farms). Such patterns in terms of distribution of local crop diversity are quite common in the literature and consistent with the findings of [Jarvis et al. \(2008\)](#), who found that growing area and landrace diversity are related, and similar to those of [Zimmerer \(1991\)](#) for the distribution of potato biodiversity in Andean Peru.

From an ethnobiological point of view, our findings reflect the differing ways of managing specific (crop species) and infra-specific crop diversity (landraces). Growing numerous species is more complicated than growing numerous landraces, for several reasons. First, each species has its specific needs in terms of soil quality and preparation, sowing date, quantity of labour required and when it is required ([Gariné and Raimond, 2005](#)). Among landraces of the same species, these needs are not so divergent. Farmers possessing a relatively large land area have more chance to encounter different soil types and quality among their fields. Also, larger farms or those with an extensive social network can expect to have an adequate labour supply ([Abizaid et al., 2015](#)) to grow a large portfolio of species ([Gariné and Raimond, 2005](#)). Thus, farmers with more assets, including social capital and labour, tend to cultivate larger and more numerous fields and have greater crop diversity ([Alvarez et al., 2005](#); [Coomes and Ban, 2004](#); [Zimmerer, 1991](#)). Smallholder poverty may limit the diversity of crops that can be raised. Previous studies concluded that certain species are needed to meet basic needs (e.g. food, medicinal, etc.) and other species are more optional, reflected by higher levels of infra-specific diversity for staples compared to other crops ([Jarvis et al., 2008](#)), especially under stressful abiotic conditions ([Labeysrie et al., 2013](#)). Another possible explanation of the lower

heterogeneity for degrees for landraces is that several landraces of the main species may be grown to fill diverse needs driven by cultural and dietary preferences, shifts in market demand and labour availability (Brush and Meng, 1998; Gauchan et al., 2005; Johns et al., 2013), heterogeneity in soil and water resources (Bellon and Taylor, 1993; Bisht et al., 2007), biotic stresses (Finckh and Wolfe, 2006) and the need to enhance pollination levels via outcrossing (Kremen et al., 2002). Much infra-specific diversity is held at the community level rather than within individual farms (Brush et al., 2015; Mulumba et al., 2012). In addition, in agroecosystems where many species are grown, farms maintaining collections of landraces will be few because less varietal diversity of the crop species is available to the farmer due to financial, social or policy constraints. Finally, the reason for a greater heterogeneity of crop diversity at the specific level compared to the infra-specific level may lie in the traits of the crop species considered in the analysis and their reproductive systems. In their broad comparison of nomenclature systems, Jarvis et al. (2008) showed that farmers use more detailed classifications for clonally reproduced crops than for self-fertilizing, partially self-fertilizing or outcrossing crops. This pattern was confirmed in our dataset. The only cases where over-distribution of farm degree was observed at the infra-specific level (ME-M01, SC-M04 and SC-M07) were all villages in which the staple food was provided by clonally propagated species (manioc and taro).

We applied additional tests to detect more detailed patterns in crop diversity within the meta-dataset and the sources of divergence in terms of crop portfolio composition. Our analysis of outliers identified certain farms that held unique portfolios of species or landraces. In most cases, it is the high-diversity farms that mainly contribute to the global diversity. These findings are consistent with the hypothesis of nestedness and of sink–source dynamics described in Alvarez et al. (2005) and Coomes (2010), and frequently postulated to be important, in the dynamics of local diversity, of one or a small number of experts or nodal farmers in a village (Boster, 1983; Kawa et al., 2013; Padoch and Jong, 1991; Peroni and Hanazaki, 2002; Perrault-Archambault and Coomes, 2008; Salick et al., 1997; Subedi et al., 2003; Tapia, 2000).

Nevertheless, it would be incorrect to say that this is a consistent tendency in the meta-dataset. Indeed, we observed the opposite relationship in other datasets whereby low-diversity farms contributed significantly to the local diversity (EG-M05, EG-M08, DJ-M015a and SC-M07). In some cases, one or a few farmers grew rarer species or landraces due to curiosity,

for aesthetic reasons, or to maintain a social status of expert at the local level (Elias et al., 2000; Hawkes, 1983; Meilleur, 1998), or to have an object that others do not have (Coomes and Ban, 2004). Possessing a rare species or landrace might, for instance, allow a young farmer to distinguish himself from others to develop niche market (rare vegetables and tobacco) or for other non-economic reasons. Possessing an object that others do not could increase its potential transfer value to other members of the community (Caillon and Lanouguère-Bruneau, 2005). Additional factors influence the distribution of local crop diversity, for instance, the role played by differences associated with gender and generation, access to seed markets, farmers' food preferences and the market value of crops. Patterns of vertical transmission of seeds from mother-in-law to daughter-in-law (Delêtre et al., 2011; Labeyrie et al., 2013) or from father/mother-in-law to son-in-law (Wencélius and Garine, 2015) in patrilineal societies with virilocal rules of residence, i.e., where the son and his wife (wives) stay in the same village of the son's father, generation after generation, may constitute another source of divergence in crop diversity among families from the same village.

Considering now the village unit as a complex system, patterns of crop distribution at both the species and landrace levels are shaped by the self-organized action of the farmers, resulting from the sum of individual choices. This behaviour can be interpreted as a 'collective knowledge' that maintains crop diversity, to cope with multiple environmental and socio-cultural constraints and perturbations, and to maintain cultural cohesion through seed circulation (Empereire and Peroni, 2007). These self-organized distributions of crop diversity are vulnerable, depending on their pattern and the type of perturbation. For instance, maintenance of crop diversity may be threatened if local crop diversity is concentrated in a few crop-rich farms, should a disaster happen. Local farmer populations can be expected to be more vulnerable to outbreaks and rapid spread of pests or pathogens when crop-rich and crop-poor farms from the same village both grow common species (used as staple food) or common landraces. Therefore, cultivating both common and rare landraces on the same farm increases farms' resilience in case of major pests affecting the most common landraces.

These multiple patterns of crop diversity raise particular concern about the issues around the conservation of crop diversity. By detecting how local diversity is distributed, our methods could help scientists involved in *ex situ* and *in situ* conservation programs to optimize their sampling strategies for plant collection and farmers involvement, respectively. In addition to the statistical methods developed in this chapter, LBM and PCA are visualizations

derived from network data, and may serve as useful tools in communicate information about the distribution of crop diversity at the local scale with NGOs, politicians, farmers and all the stakeholders interested by crop management, as suggested by [Pocock et al. \(2015\)](#).

More generally, because these distinct patterns of crop diversity have been detected in different agro-ecological environments and socio-cultural contexts without controlling for other potential factors (and without additional information about each village), it is not yet possible to assess how one particular agro-ecological environment and socio-cultural context shapes the distribution of local crop diversity. Additional studies are needed in this direction to detect the local drivers influencing the observed distribution of crop diversity by collecting data to characterize specific and infra-specific diversity of crop plants and socio-cultural diversity of farmers. Such investigations will help us in understanding trade-offs between ecological and socio-cultural functions within agroecosystems.

## 5.2 Relevance of Network-Based Methods

The network-based methods introduced in this chapter provide a set of useful tools to analyse the distribution of local diversity in crop species and varieties. Indeed, our framework allowed us to answer four key questions:

1. Are farm and crop degrees more variable than expected under a null model which assumes a homogeneous probability of interaction between potential partners?
2. Are crop-by-farm interactions structured by blocks and, if so, what are the characteristics of these blocks?
3. Are certain crops or certain farms obvious outliers in their pattern of interactions?
4. Do crop-poor (low-degree) and crop-rich (high-degree) farms contribute significantly more or less than expected, based solely on knowledge of their crop-richness (degrees), to the overall diversity of crops cultivated locally?

By combining these different indices, tests and metrics, we provide a realistic and complete picture of the complex structure of crop diversity. This framework readily detected cases, for example, in which crop diversity is different in two different villages (through the LBMs) and identified farms—be they low-degree or high-degree farms—as unique and important providers of crop diversity (through uncovering of outliers in PCA and measures of uniqueness).

One strength of this framework is the use of a hierarchy of null models of increasing complexity. For instance, the simplest model for a bipartite network with variable degrees is the Erdős-Rényi  $G(N, p)$  model in which interaction between nodes from the two different categories is restricted (each link has the same probability of occurring). Deviations from this null model allow assessment of degree heterogeneity or the presence of blocks (groups of farms that preferentially cultivate a certain group of species). When looking for more elaborate structures in the network (and not only in degree distributions), we relied on the configuration model, which randomizes interactions while keeping all degrees in the network constant. Consequently, one can disentangle whether the observed patterns, such as the block structure, are simply explained by the degree heterogeneity or are truly emergent properties. Furthermore, the above approaches (LBMs and PCA) provide visualization methods of the network highlighting its different characteristics, e.g., modules or outliers. These graphical representations are complementary to the more usual network representations reviewed in [Pocock et al. \(2015\)](#). It is important to note that our network-based approach can foster transdisciplinarity as it can be extended to datasets from other disciplines, including ecology, to detect particular patterns in bipartite networks ([Mulder et al., 2015](#)), especially with the outcome of next-generation sequencing techniques ([Vacher et al., 2015](#)). In ecology, the tests could efficiently supplement metrics that are routinely used, such as modularity or nestedness scores ([Fortuna et al., 2010](#)). Depending on the size of the dataset, LBMs can be as informative as traditional modularity-computing techniques (or even more informative) in finding underlying structures within bipartite datasets ([Leger, 2015](#)). Moreover, LBMs can also elucidate non-modular blocks, such as quasi-partite structures (i.e. when such structures are not exactly bi- or multi-partite but quite close to one of those) within a network. Of course, the power of all such methods depends heavily on the number of nodes in the network, but the application to ecological questions of the set of methods proposed here could readily generate much more informative descriptions of ecological networks than connectance, modularity and nestedness scores alone.

The approach used in this chapter does not rely on a direct estimation of nestedness, because the different methods available to compute nestedness do not converge ([Fig. A6](#)). However, the set of methods designed here to uncover the uniqueness of contributions to diversity of crop-rich and crop-poor farms actually provide complementary information on whether specialists interact preferentially with generalists, as assumed under a ‘nested’

scenario in ecology, or not. We thus suggest that this toolkit could be used as an alternative to the classical methods for detecting nestedness that are usually applied to ecological datasets (Podani and Schmera, 2012). For future use, the code is available at the following URL: <http://netseed.cesab.org/>.

From a methodological point of view, the configuration model must be accompanied by several caveats. Most prominently, the fact that the degrees of all nodes are constant makes the model highly constrained. Chung and Lu (2002a,b) developed a model that generated graphs with given expected degrees, relaxing the requirement that all samples of the model reproduce exactly the observed degrees. Degrees of networks sampled from this model are allowed to vary slightly around a fixed expected value. Interestingly, the Chung-Lu model has recently been extended into the so-called degree-corrected stochastic block model (Karrer and Newman, 2011) incorporating both degree-heterogeneity parameters as in the Chung-Lu model and a block structure as in the LBM. Such models would allow disentangling the farms' overall crop richness, as well as crop rarity, from the preferences of certain farms for specific groups of crops (block structure). Inference methods for this model have been developed recently (e.g. Lei and Rinaldo, 2014). However, the complexity of these models makes the estimation (and the computation of  $p$ -values) unreliable for small networks such as those considered in this study. Nevertheless, the Chung-Lu model and degree-correcting stochastic block models are promising directions for research on larger-scale ecological networks.



---

## 6. CONCLUSION

In this chapter, we develop new network-based indicators and statistical tests to characterize patterns of crop diversity at local scales. We applied this methodological framework to a meta-dataset from 10 countries containing inventory data at the specific or infra-specific level. Our results identify different sources of heterogeneity in local crop diversity:

- i. diversity at the specific level is generally much more heterogeneous among farms compared to diversity at the infra-specific level;
- ii. two or more groups of farms can be identified based on their unique crop richness and



- iii. although diversity-rich farms often contribute most to global diversity, in some cases diversity-poor farms contribute equally with rare species and varieties.

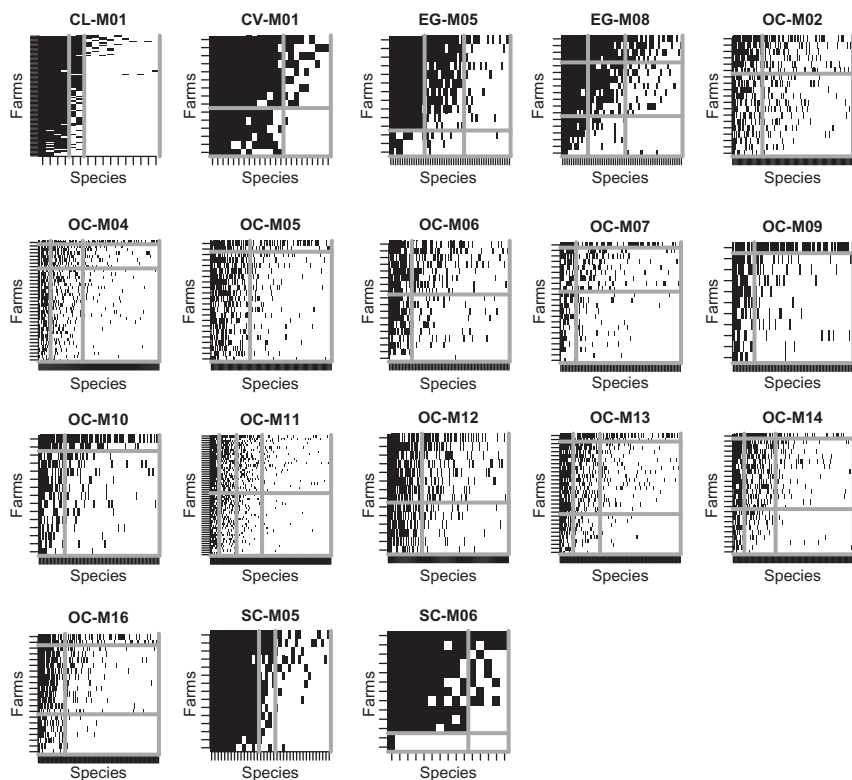
This analysis reveals the absence of any general pattern of crop diversity distribution at the village level, indicating a strong dependence on agro-ecological and socio-cultural contexts. These results suggest that local communities adapt self-organized strategies to their growing contexts. Further empirical investigations are needed to disentangle the different drivers shaping crop diversity distribution, more particularly comparing the impacts of biological properties of crops (open-pollinated vs. self-pollinated crop, seed vs. cuttings, annual vs. perennial, etc.), of social organization of farmers (patrilinearity vs. matrilinearity, local community vs. community of practices), of agricultural policy and of diversity of ecological landscapes (open vs. closed systems). Our methodological framework provides a useful approach and an informative overview of patterns in the distribution of diversity. The toolkit developed and applied in this study offers an alternative approach to the classical methods of detecting nestedness, in both ethnographic and ecological datasets. More broadly, this methodological framework—which helps to detect patterns of crop distribution within local social organizations—enables the investigation of trade-offs between ecological and social functions of agroecosystems within a same analytical framework.

## ACKNOWLEDGEMENTS

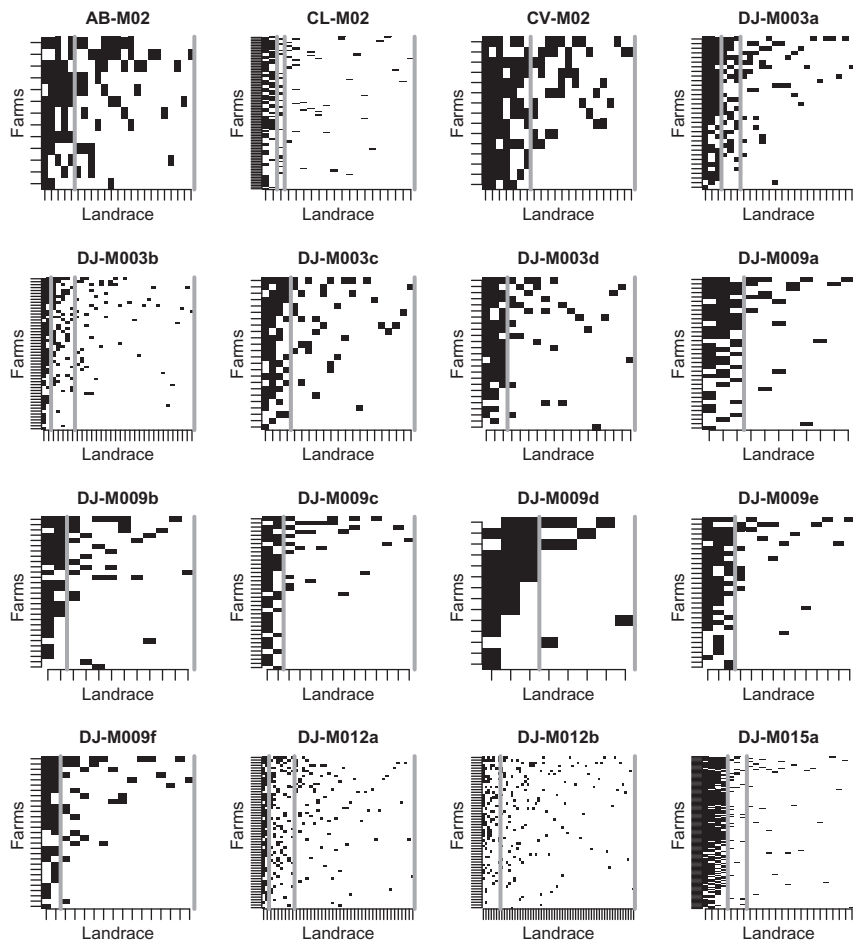
The authors gratefully acknowledge support from France's Foundation pour la Recherche sur la Biodiversité (FRB) that made possible NetSeed, an international collaboration of researchers studying farmer seed networks. Mathieu Thomas was provided with a six-month post-doctoral fellowship within this program. The Centre de Synthèse et d'Analyse sur la Biodiversité (CESAB) provided essential logistical support for regular workshops on the subject, enabling us to develop and mature both the theoretical and methodological aspects of our research. Additional support of the ongoing research collaboration through the MIREs and MADRES networks was provided by the following agencies: the Réseau National des Systèmes Complexes (RNSC), the Institut National de la Recherche Agronomique (INRA) and the Centre National de la Recherche Scientifique (CNRS). We are most grateful to the NETSEED researchers not directly included in the preparation of this chapter for the fruitful discussions in which they participated during the project's different meetings.



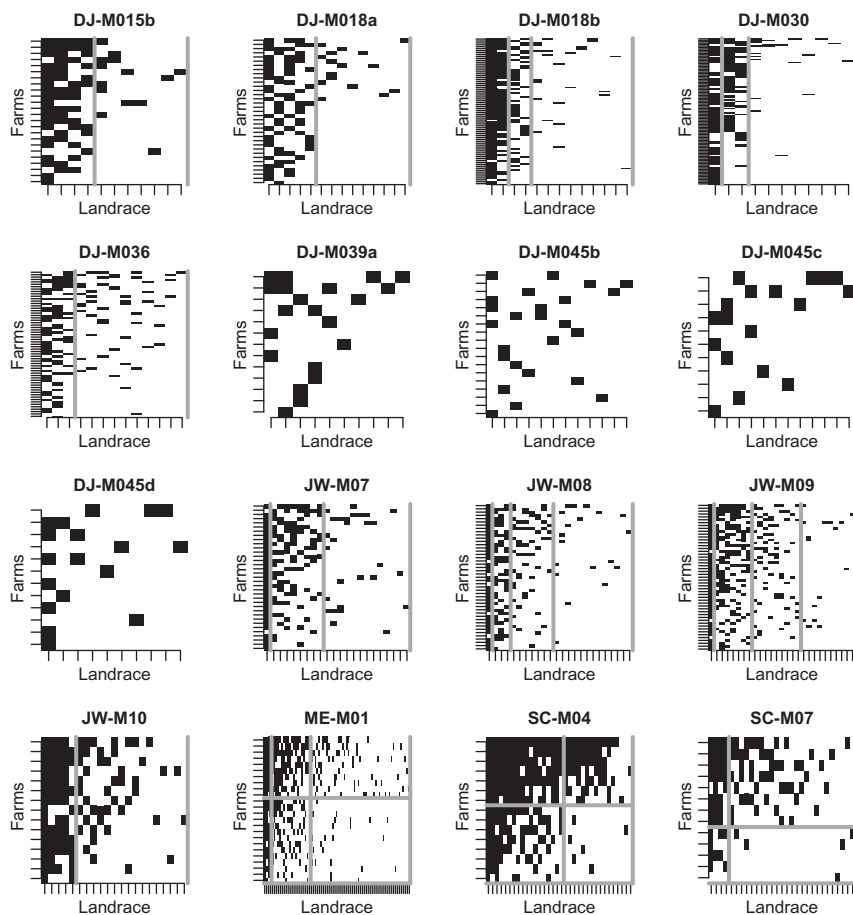
## APPENDIX. LBM Representation



**Figure A1** Representation of the incidence matrix for the 18 datasets collected at the specific level. The left panel corresponds to the original matrix without reordering, the right panel corresponds to the reordering based on block detection using the LBM method and density of the graph. The higher density is always on the top left side of the matrix.

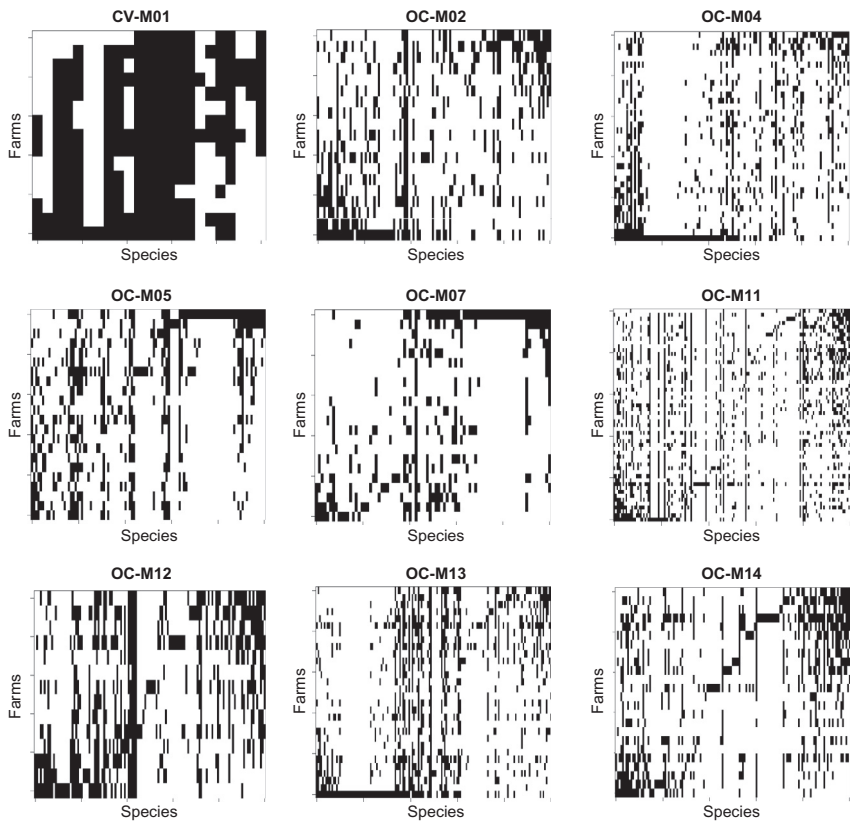


**Figure A2** Representation of the incidence matrix for the 32 datasets collected at the infra-specific level. The left panel corresponds to the original matrix without reordering, the right panel corresponds to the reordering based on block detection using the LBM method and density of the graph. The higher density is always on the top left side of the matrix. Part 1.

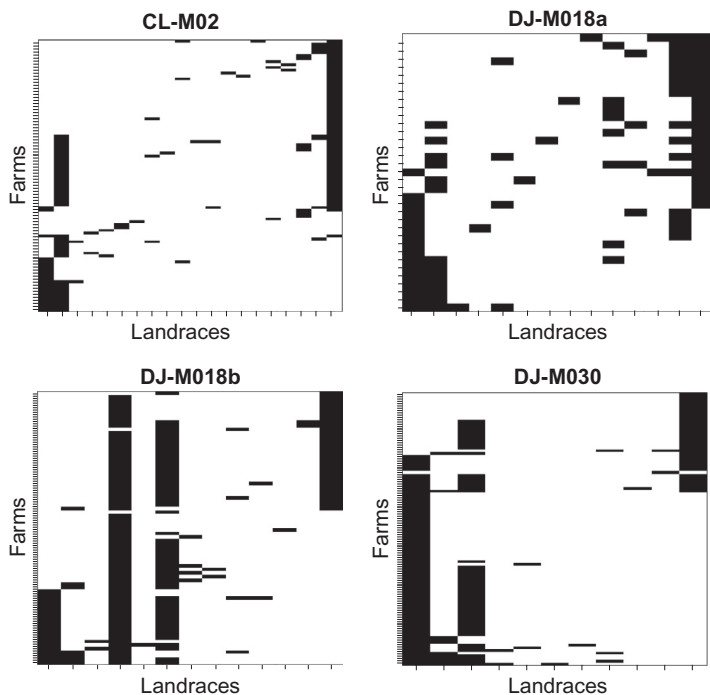


**Figure A3** Representation of the incidence matrix for the 32 datasets collected at the infra-specific level. The left panel corresponds to the original matrix without reordering, the right panel corresponds to the reordering based on block detection using the LBM method and density of the graph. The higher density is always on the top left side of the matrix. Part 2.

## Outlier Representation



**Figure A4** Representation of the PCA residuals on the nine datasets that yielded significant results at the specific level.



**Figure A5** Representation of the PCA residuals on the four datasets that gave significant results at the infra-specific level.

### Statistical Power Study of the Test Measuring the Impact of Crop-Rich and Crop-Poor Farms

The same model as in [Section 3.5.4](#) is used for studying the behaviour of this test, introduced in [Section 3.5.2](#). The three different settings of parameters correspond to an edge density of approximately 0.18. Thousand incidence matrices were simulated in each of the three settings with different incidence matrix sizes:  $n=20$ , 50 and  $m=20$ , 50.

**Table A1** Estimated Probabilities of Rejection of the Null Hypothesis (in %) for the Different Contribution Tests Under the Three Toy Models for Two Alpha Levels, 1% and 5%, When: (a)  $n = 50, m = 50$ , (b)  $n = 50, m = 20$ , (c)  $n = 20, m = 50$  and (d)  $n = 20, m = 20$

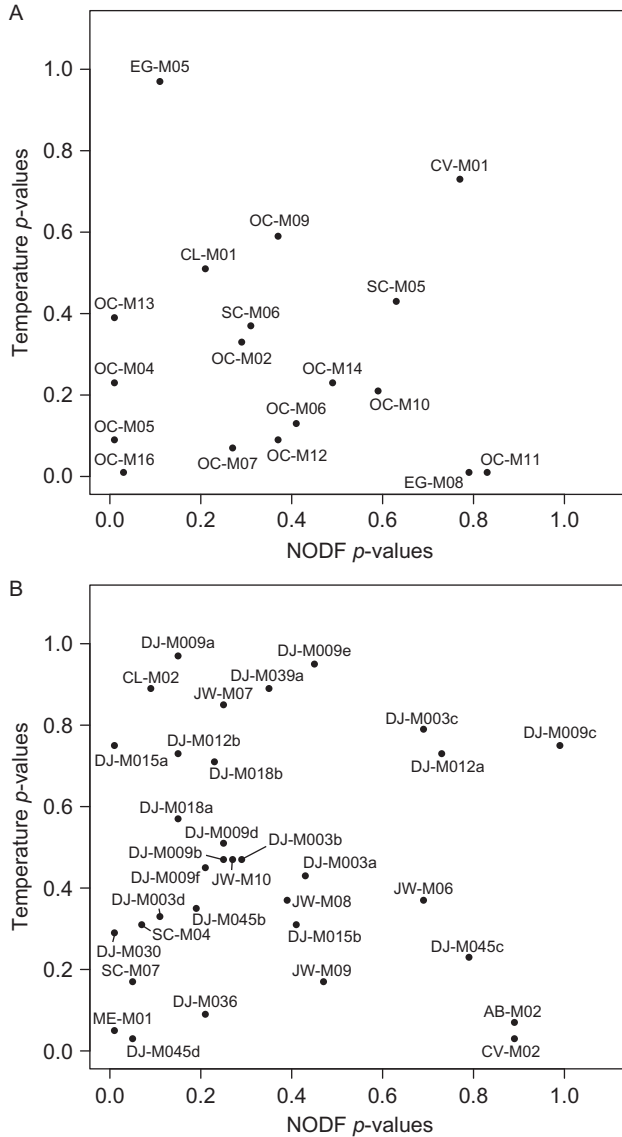
Alpha Level	Fig. 12		Fig. 13		Fig. 14	
	1.00%	5.00%	1.00%	5.00%	1.00%	5.00%
<b>(a)</b>						
E.diff.pvalue.L	0.3	4.1	71	89.5	0.1	0.3
E.diff.pvalue.R	0.4	5.5	0	0	41	64.2
Hbeta.rich.pvalue.L	0.3	4.2	79.2	94.7	0.1	0.3
Hbeta.rich.pvalue.R	0.7	4.9	0	0	39.7	63.6
Hbeta.poor.pvalue.L	0.7	4.9	0	0	39.7	63.6
Hbeta.poor.pvalue.R	0.3	4.2	79.2	94.7	0.1	0.3
<b>(b)</b>						
E.diff.pvalue.L	0.9	4.9	17.2	38.2	0.09	2.6
E.diff.pvalue.R	0.1	5.5	0	0	4.7	14.1
Hbeta.rich.pvalue.L	1.3	4.7	20.3	43.3	0.9	3.1
Hbeta.rich.pvalue.R	0.5	5.3	0	0	5.3	14.8
Hbeta.poor.pvalue.L	0.5	5.3	0	0	5.3	14.8
Hbeta.poor.pvalue.R	1.3	4.7	20.3	43.3	0.9	3.1
<b>(c)</b>						
E.diff.pvalue.L	0.9	6.1	16.8	40	0.3	0.5
E.diff.pvalue.R	1.4	5.7	0	0.2	12.3	29.8
Hbeta.rich.pvalue.L	1.3	6	26.7	52.5	0.3	0.6
Hbeta.rich.pvalue.R	1.1	5.6	0	0	12.3	28.9
Hbeta.poor.pvalue.L	1.1	5.6	0	0	12.3	28.9
Hbeta.poor.pvalue.R	1.3	6	26.7	52.5	0.3	0.6
<b>(d)</b>						
E.diff.pvalue.L	0.7	4.6	5.5	19.1	0.9	3.8
E.diff.pvalue.R	1.1	4.6	0	1.4	2.5	7.7
Hbeta.rich.pvalue.L	0.8	4.6	8.4	22.8	0.7	3.8
Hbeta.rich.pvalue.R	0.8	5.5	0	1.3	2.1	7.4
Hbeta.poor.pvalue.L	0.8	5.3	0	1.3	2.1	7.4
Hbeta.poor.pvalue.R	0.8	4.6	8.5	23.2	0.7	3.8

Table A1 depicts the proportion of rejection (in %) as a function of incidence matrix size when the  $\alpha$ -level is set to 1% and 5%. In Settings 2 and 3 (alternative hypothesis), the rejection probability exhibits the same pattern. When there are only  $n=20$  farms or  $m=20$  crops, the power is quite low, whereas for larger matrices ( $n=m=50$ ), the power is greatly increased. Under the first setting without interaction between richness of the farms and the status of crops, the  $p$ -values are nearly uniformly distributed on  $[0,1]$ . These simulations confirm that our test is able to detect contrasted contribution to the diversity by ‘crop-rich’ and ‘crop-poor’ farms as long as the sample size is large enough.

### Estimation of Nestedness

This section describes the nestedness results obtained on the meta-dataset using two methods: the temperature (Rodríguez-Gironés and Santamaría, 2006) and the NODF (Almeida-Neto et al., 2008). Figure A6 represents the  $p$ -values computed for each estimator after re-sampling using the configuration model introduced in Section 3.4.1. Our results are consistent with those of Podani and Schmera (2012), because for the same meta-dataset, tests performed with one or the other index were inconsistent.





**Figure A6** Plot representing on the x-axis the NODF  $p$ -values computed by re-sampling and on the y-axis the temperature  $p$ -values computed by re-sampling. In both cases, re-sampling was performed using the configuration model: (a) for datasets collected at the specific level and (b) for datasets collected at the infra-specific level.

## GLOSSARY

**Network** a set of interconnected actors (human or non-human) and formally modelled by a graph.

**Bipartite network** a network whose nodes can be partitioned into two disjoint subsets ( $F$  to represent the farm and  $C$  to represent the crop: species/landraces) such that no edge connects two nodes from  $F$  or two nodes from  $C$ .

**Node** synonymous with ‘vertex’. A node is the fundamental unit of which graphs are formed.

**Edge** an edge is a link between two nodes. Every edge has two endpoints in the sets of nodes. In the particular case of bipartite networks, the two endpoints belong to two disjoint subsets of nodes, e.g., farms ( $F$ ) and crops ( $C$ , species or landraces). The presence of an edge indicates that the considered crop is cultivated on the considered farm.

**Interaction network** a network of nodes that are connected by features. In a crop-by-farm interaction network, crops are cultivated by farmers who are members of the farm.

**Nestedness** this concept, for which different indices have been devised, aims at quantifying the extent to which nodes of one subset (e.g.  $F$ ) with low degrees are linked to nodes of the other subset (e.g.  $C$ ) with high degrees. In the example of crop-by-farm networks, indices of nestedness aim to measure to what extent ‘crop-poor’ farms grow a subset of the crops cultivated on ‘crop-rich’ farms.

**Degree** the number of edges incident to a vertex. A farm’s degree is the number of crops cultivated on the considered farm.

**Configuration model** a random graph model with a prescribed degree sequence. All graphs with this degree sequence obtained by permutation are equiprobable in this model (for details, see [Section 3.4.1](#)).

**Graph** a mathematical concept defined by a finite set of nodes (vertices) connected by edges (links).

**Random graph model** a generative model of graphs where the set of nodes is deterministic and the edges are drawn according to some probability distribution.

**Erdős-Rényi model** a random graph model in which all the edges are drawn independently with the same probability  $p$ .

**Latent-block model** a random graph model that assumes that the nodes belong to (unobserved) blocks and that the probability of connection between two nodes depends only on the blocks they belong to. This block structure can be estimated, allowing the clustering of nodes (farms or crops) based on similarities in terms of connectivity properties (see [Section 3.3](#)).

**Incidence matrix** 0/1 matrix  $\mathbf{A}$ . Its rows are indexed by the set of farms  $F$  and its columns are indexed by the set of crops  $C$ . The entry  $A_{ij}$  equals one if and only if crop  $j$  is cultivated by farmers on farm  $i$  (see [Section 3.1](#)).

## REFERENCES

- Abizaid, C., Coomes, O.T., Takasaki, Y., Brisson, S., 2015. Social network analysis of peasant agriculture: cooperative labor as gendered relational networks. *Prof. Geogr.* 67, 447–463.
- Almeida-Neto, M., Guimarães, P., Guimarães, P.R., Loyola, R.D., Ulrich, W., 2008. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* 117, 1227–1239.

- Alvarez, N., Garine, E., Khasah, C., Dounias, E., Hossaert-McKey, M., McKey, D., 2005. Farmers' practices, metapopulation dynamics, and conservation of agricultural biodiversity on-farm: a case study of sorghum among the Duupa in sub-Saharan Cameroon. *Biol. Conserv.* 121, 533–543.
- Astegiano, J., Guimarães Jr., P.R., Cheptou, P.-O., Vidal-Morais, M.M., Mandai, C.Y., Ashworth, L., Massol, F., 2015. Persistence of plant-pollinator networks in spite of habitat loss: insights from trait-based metacommunity models. *Adv. Ecol. Res.* 53.
- Atmar, W., Patterson, B.D., 1993. The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia* 96, 373–382.
- Bascompte, J., Jordano, P., 2007. Plant-animal mutualistic networks: the architecture of biodiversity. *Annu. Rev. Ecol. Syst.* 38, 567–593.
- Bastolla, U., Fortuna, M.A., Pascual-García, A., Ferrera, A., Luque, B., Bascompte, J., 2009. The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* 458, 1018–1020.
- Bellon, M.R., Taylor, J.E., 1993. 'Folk' soil taxonomy and the partial adoption of new seed varieties. *Econ. Dev. Cult. Chang.* 41, 763–786.
- Bianchi, F., Booij, C.J.H., Tscharntke, T., 2006. Sustainable pest regulation in agricultural landscapes: a review on landscape composition, biodiversity and natural pest control. *Proc. R. Soc. B Biol. Sci.* 273, 1715–1727.
- Bickel, P.J., Sarkar, P., 2015. Hypothesis testing for automated community detection in networks. *J. Roy. Stat. Soc. B Stat. Meth.* (early view).
- Bisht, I., Mehta, P., Bhandari, D., 2007. Traditional crop diversity and its conservation on-farm for sustainable agricultural production in Kumaon Himalaya of Uttaranchal State: a case study. *Genet. Resour. Crop. Evol.* 54, 345–357.
- Bonneuil, C., Goffaux, R., Bonnin, I., Montalent, P., Hamon, C., Balfourier, F., Goldinger, I., 2012. A new integrative indicator to assess crop genetic diversity. *Ecol. Indic.* 23, 280–289.
- Boster, J., 1983. A comparison of the diversity of Jivaroan gardens with that of the tropical forest. *Hum. Ecol.* 11, 47–68.
- Boster, J.S., 1985. Selection for perceptual distinctiveness: evidence from Aguaruna cultivars of *Manihot esculenta*. *Econ. Bot.* 39, 310–325.
- Brush, S.B., Meng, E., 1998. Farmers' valuation and conservation of crop genetic resources. *Genet. Resour. Crop. Evol.* 45, 139–150.
- Brush, S.B., Bellon, M.R., Hijmans, R.J., Ramirez, Q.O., Perales, H.R., Etten, J.v., 2015. Assessing maize genetic erosion. *Proc. Natl. Acad. Sci. U.S.A.* 112, E1.
- Butterfield, et al., 2015. Tradeoffs and compatibilities among ecosystem services: biological, physical and economic drivers of multifunctionality. *Adv. Ecol. Res.* 53.
- Caillon, S., Lanouguère-Bruneau, V., 2005. Gestion de l'agrobiodiversité dans un village de Vanua Lava (Vanuatu): stratégies de sélection et enjeux sociaux. *J. Soc. Océanistes* 120–121, 129–148.
- Chung, F., Lu, L., 2002a. The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15879–15882.
- Chung, F., Lu, L., 2002b. Connected components in random graphs with given expected degree sequences. *Ann. Comb.* 6, 125–145.
- Connor, E.F., Simberloff, D., 1979. The assembly of species communities: chance or competition? *Ecology* 60, 1132–1140.
- Coomes, O.T., 2010. Of stakes, stems, and cuttings: the importance of local seed systems in traditional Amazonian societies. *Prof. Geogr.* 62, 323–334.
- Coomes, O.T., Ban, N., 2004. Cultivated plant species diversity in home gardens of an Amazonian peasant village in Northeastern Peru. *Econ. Bot.* 58, 420–434.
- Crowder, D.W., Northfield, T.D., Strand, M.R., Snyder, W.E., 2010. Organic agriculture promotes evenness and natural pest control. *Nature* 466, 109–112.

- Delètre, M., McKey, D.B., Hodkinson, T., 2011. Marriage exchanges, seed exchanges, and the dynamics of manioc diversity. *Proc. Natl. Acad. Sci. U.S.A.* 108, 18249–18254.
- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* 39, 1–38.
- Di Falco, S., Perrings, C., 2005. Crop biodiversity, risk management and the implications of agricultural assistance. *Ecol. Econ.* 55, 459–466.
- Diamond, J., 2002. Evolution, consequences and future of plant and animal domestication. *Nature* 418, 700–707.
- Drinkwater, L.E., Wagoner, P., Sarrantonio, M., 1998. Legume-based cropping systems have reduced carbon and nitrogen losses. *Nature* 396, 262–265.
- Elias, M., Rival, L., McKey, D., 2000. Perception and management of cassava (*Manihot esculenta* Crantz) diversity among Makushi Amerindians of Guyana (South America). *J. Ethnobiol.* 20, 239–265.
- Emperaire, L., Peroni, N., 2007. Traditional management of agrobiodiversity in Brazil: a case study of manioc. *Hum. Ecol.* 35, 761–768.
- Finckh, M.R., Wolfe, M.S., 2006. Diversification strategies. In: Cooke, B.M., Jones, D., Gareth, B.K. (Eds.), *The Epidemiology of Plant Diseases*. Springer, Netherlands, pp. 269–307.
- Fortuna, M.A., Stouffer, D.B., Olesen, J.M., Jordano, P., Mouillot, D., Krasnov, B.R., Poulin, R., Bascompte, J., 2010. Nestedness versus modularity in ecological networks: two sides of the same coin? *J. Anim. Ecol.* 79, 811–817.
- Fraser, J., Alves-Pereira, A., Junqueira, A., Peroni, N., Clement, C., 2012. Convergent adaptations: bitter manioc cultivation systems in fertile anthropogenic dark earths and flood-plain soils in Central Amazonia. *PLoS One* 7, e43636.
- Garine, E., Raimond, C., 2005. La culture intensive fait-elle disparaître la biodiversité? In: *Dynamique de la Biodiversité et Modalité d'Accès aux Milieux et aux Ressources*. Institut Français de la Biodiversité, Paris, France, pp. 25–28.
- Gauchan, D., Smale, M., Chaudhary, P., 2005. Market-based incentives for conserving diversity on farms: the case of rice landraces in Central Tarai, Nepal. *Genet. Resour. Crop. Evol.* 52, 293–303.
- Gil, et al., 2015. Protecting an ecosystem service: approaches to understanding and mitigating threats to wild insect pollinators. *Adv. Ecol. Res.* 53.
- Govaert, G., Nadif, M., 2008. Block clustering with Bernoulli mixture models: comparison of different approaches. *Comput. Stat. Data Anal.* 52, 3233–3245.
- Hawkes, J.G., 1983. *The Diversity of Crop Plants*. Harvard University Press, London, UK.
- Heckler, S., Zent, S., 2008. Piara manioc varietals: hyperdiversity or social currency. *Hum. Ecol.* 36, 679–697.
- Jackson, L.E., Pascual, U., Hodgkin, T., 2007. Utilizing and conserving agrobiodiversity in agricultural landscapes. *Agric. Ecosyst. Environ.* 121, 196–210.
- Jarvis, D.I., Brown, A.H., Cuong, P.H., Collado-Panduro, L., Latournerie-Moreno, L., Gyawali, S., Tanto, T., Sawadogo, M., Mar, I., Sadiki, M., 2008. A global perspective of the richness and evenness of traditional crop-variety diversity maintained by farming communities. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5326–5331.
- Johns, T., Powell, B., Maundu, P., Eyzaguirre, P.B., 2013. Agricultural biodiversity as a link between traditional food systems and contemporary development, social integrity and ecological health. *J. Sci. Food Agric.* 93, 3433–3442.
- Jordano, P., 1987. Patterns of mutualistic interactions in pollination and seed dispersal: connectance, dependence asymmetries, and coevolution. *Am. Nat.* 129, 657–677.
- Jordano, P., Bascompte, J., Olesen, J.M., 2003. Invariant properties in coevolutionary networks of plant–animal interactions. *Ecol. Lett.* 6, 69–81.
- Karrer, B., Newman, M.E., 2011. Stochastic blockmodels and community structure in networks. *Phys. Rev. E* 83, 016107.

- Kawa, N.C., McCarty, C., Clement, C.R., 2013. Manioc varietal diversity, social networks, and distribution constraints in rural Amazonia. *Curr. Anthropol.* 54, 764–770.
- Keribin, C., Brault, V., Celeux, G., Govaert, G., 2014. Estimation and selection for the latent block model on categorical data. *Stat. Comput.* 25, 1–16.
- Kolaczyk, E.D., 2009. *Statistical Analysis of Network Data: Methods and Models*. Springer Science & Business Media, New York, USA.
- Kremen, C., Williams, N.M., Thorp, R.W., 2002. Crop pollination from native bees at risk from agricultural intensification. *Proc. Natl. Acad. Sci. U.S.A.* 99, 16812–16816.
- Labeyrie, V., Rono, B., Leclerc, C., 2013. How social organization shapes crop diversity: an ecological anthropology approach among Tharaka farmers of Mount Kenya. *Agric. Hum. Values* 31, 97–107.
- Lande, R., 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos* 76, 5–13.
- Lazega, E., Mounier, L., Snijders, T., Tubaro, P., 2012. Norms, status and the dynamics of advice networks: a case study. *Soc. Networks* 34, 323–332.
- Leclerc, C., Coppens d'Éeckenbrugge, G., 2012. Social organization of crop genetic diversity. The  $G \times E \times S$  interaction model. *Diversity* 4, 1–32.
- Leger, J.-B., 2015. Blockmodels: Latent and Stochastic Block Model Estimation by a 'V-EM' Algorithm. In: R package version 1.1.1.
- Lei, J., Rinaldo, A., 2014. Consistency of spectral clustering in stochastic block models. *Ann. Stat.* 43, 215–237.
- Macfadyen, S., Bohan, D.A., 2010. Crop domestication and the disruption of species interactions. *Basic Appl. Ecol.* 11, 116–125.
- Mariac, C., Jehin, L., Saidou, A.-A.A., Vigouroux, Y., 2011. Genetic basis of pearl millet adaptation along an environmental gradient investigated by a combination of genome scan and association mapping. *Mol. Ecol.* 20, 80–91.
- Martin, J.F., Roy, E.E.D., Diemont, S., Ferguson, B.G., 2010. Traditional ecological knowledge (TEK): ideas, inspiration, and designs for ecological engineering. *Ecol. Eng.* 36, 839–849.
- Meilleur, B.A., 1998. Clones within clones: cosmology and esthetics and Polynesian crop selection. *Anthropologica* 40, 71–82.
- Miklós, I., Podani, J., 2004. Randomization of presence–absence matrices: comments and new algorithms. *Ecology* 85, 86–92.
- Millennium Ecosystem Assessment (MEA), 2005. *Ecosystems and Human Well-Being: Biodiversity Synthesis*. World Resources Institute, Washington, DC, USA.
- Mulder, et al., 2015. 10 years later: revisiting priorities for science and society a decade after the millennium ecosystem assessment. *Adv. Ecol. Res.* 53.
- Mulumba, J.W., Nankya, R., Adokorach, J., Kiwuka, C., Fadda, C., De Santis, P., Jarvis, D.I., 2012. A risk-minimizing argument for traditional crop varietal diversity use to reduce pest and disease damage in agricultural ecosystems of Uganda. *Agric. Ecosyst. Environ.* 157, 70–86.
- Padoch, C., Jong, W.d., 1991. The house gardens of Santa Rosa: diversity and variability in an Amazonian agricultural system. *Econ. Bot.* 45, 166–175.
- Peroni, N., Hanazaki, N., 2002. Current and lost diversity of cultivated varieties, especially cassava, under swidden cultivation systems in the Brazilian Atlantic Forest. *Agric. Ecosyst. Environ.* 92, 171–183.
- Perrault-Archambault, M., Coomes, O.T., 2008. Distribution of agrobiodiversity in home gardens along the Corrientes River, Peruvian Amazon. *Econ. Bot.* 62, 109–126.
- Pocock, et al., 2015. The visualisation of ecological networks, and their use as a tool for engagement, advocacy and management. *Adv. Ecol. Res.* 53.
- Podani, J., Schmera, D., 2012. A comparative evaluation of pairwise nestedness measures. *Ecography* 35, 889–900.

- Reyes-García, V., Molina, J.L., Calvet-Mir, L., Aceituno-Mata, L., Lastra, J.J., Ontillera, R., Parada, M., Pardo-de Santayana, M., Rigat, M., Vallès, J., 2013. 'Tertius gaudens': germplasm exchange networks and agroecological knowledge among home gardeners in the Iberian Peninsula. *J. Ethnobiol. Ethnomed.* 9, 1–11.
- Rival, L., McKey, D., 2008. Domestication and diversity in manioc (*Manihot esculenta* Crantz ssp. *esculenta*, Euphorbiaceae). *Curr. Anthropol.* 49, 1119–1128.
- Rodríguez-Gironés, M.A., Santamaría, L., 2006. A new algorithm to calculate the nestedness temperature of presence–absence matrices. *J. Biogeogr.* 33, 924–935.
- Salick, J., Cellinese, N., Knapp, S., 1997. Indigenous diversity of cassava: generation, maintenance, use and loss among the Amuesha, Peruvian upper Amazon. *Econ. Bot.* 51, 6–19.
- Samberg, L., Shennan, C., Zavaleta, E., 2013. Farmer seed exchange and crop diversity in a changing agricultural landscape in the southern highlands of Ethiopia. *Hum. Ecol.* 41, 477–485.
- Smale, M., Hartell, J., Heisey, P.W., Senauer, B., 1998. The contribution of genetic resources and diversity to wheat production in the Punjab of Pakistan. *Am. J. Agric. Econ.* 80, 482–493.
- Subedi, A., Chaudhary, P., Baniya, B.K., Rana, R.B., Tiwari, R.K., Rijal, D.K., Sthapit, B.R., Jarvis, D., 2003. Who maintains crop genetic diversity and how? Implications for on-farm conservation and utilization. *Cult. Agric.* 25, 41–50.
- Tapia, M.E., 2000. Mountain agrobiodiversity in Peru. *Mt. Res. Dev.* 20, 220–225.
- Thébault, E., Fontaine, C., 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* 329, 853–856.
- Tuxill, J., Reyes, L.A., Moreno, L.L., Uicab, V.C., Jarvis, D.I., 2010. All maize is not equal: maize variety choices and Mayan foodways in rural Yucatan, Mexico. In: Staller, J., Carrasco, M. (Eds.), *Pre-Columbian Foodways: Interdisciplinary Approaches to Food, Culture, and Markets in Ancient Mesoamerica*. Springer, New York, USA.
- Ulrich, W., Gotelli, N.J., 2012. A null model algorithm for presence–absence matrices based on proportional resampling. *Ecol. Model.* 244, 20–27.
- Vacher, et al., 2015. Learning ecological networks from next-generation sequencing data. *Adv. Ecol. Res.* 53.
- Vigouroux, Y., Barnaud, A., Scarcelli, N., Thuillet, A.-C., 2011. Biodiversity, evolution and adaptation of cultivated crops. *C. R. Biol.* 334, 450–457.
- Wasserman, S., Faust, K., 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, UK.
- Wencélius, J., Garine, É., 2015. Dans les sillons de l'alliance. *Ethnographie de la circulation des semences de sorgho dans l'Extrême-Nord du Cameroun*. *Les Cahiers d'Outre Mer* 265, 93–116.
- Zaman, A., Simberloff, D., 2002. Random binary matrices in biogeographical ecology— instituting a good neighbor policy. *Environ. Ecol. Stat.* 9, 405–421.
- Zimmerer, K.S., 1991. Labor shortages and crop diversity in the Southern Peruvian Sierra. *Geogr. Rev.* 81, 414–432.