Mapping fungi from below ground: online genetic resources and ectomycorrhizal geographic distributions

Tse-Laurence MA* [1], Bidartondo MI [2]

We used DNA sequences of 20 ectomycorrhizal fungal species obtained from roots in Britain and Germany to find location data within Europe for these fungi in the public DNA databases. These data were used to plot species presence on maps, environmental layers were laid over these maps, and information from those sites was extrapolated using geographic information systems. Through randomization tests the significant factors for each species from available data were tested. Similar methodology was used for fungal samples identified using morphology from the Global Biodiversity Information Facility to compare data quality and quantity. This analysis exposed the need for uniform methodology and greater distribution of sampling in order to create viable species distribution models for ectomycorrhizas.

Introduction

Ectomycorrhizal (ECM) fungi are obligate plant mutualists and they are among the most functionally important soil organisms in forest ecosystems (Smith & Read 2008). However, as the delimitation and identification of many ECM species is problematic and their life cycles largely subterranean, the geographic ranges for species are unknown. There is a need to establish current distributions in the face of changing environmental conditions, because without them even large changes in mycorrhizal distributions may go undetected.

Some ECM fungal species have conspicuous fruiting bodies that can thus be used to generate species distribution maps, *e.g.*, *Amanita phalloides* (Wolfe et al. 2010). This is often not possible as many ECM species are cryptic and difficult to observe in this fashion, *e.g.*, truffles and resupinate crusts.

☐ (1) Royal Botanic Gardens, Kew, Richmond TW9 3DS (England); (2) Imperial College London, London SW7 2AZ (England)

@ Maurice Tse-Laurence (m.tse-laurence@kew.org)

For these fungi an approach using their mycorrhizas for identification is more practical. DNA sequences of the internal transcribed spacer (ITS) region of the nuclear ribosomal DNA provide a universal genetic marker for fungi. This study makes use of their growing availability in online DNA databases to obtain spatial presence data for ECM species thus far unmapped.

Ryberg et al. (2008) studied the strength of GenBank for meta-analysis and identification of ECM fungi with a focus on illustrating the gaps in identification for the genus *Inocybe*, but they also analysed the location of fungal species from GenBank providing a rough idea of their distribution on a whole-country basis. This was an early example demonstrating the potential for a DNA sequence method for mycorrhizal mapping. Two recent studies have applied spatial data on fungal presence to generate Species Distribution Models (SDM). Wollan et al. (2008) used *herbarium* mushroom records to create a fungal SDM for Norway, and Wolfe et al. (2010) gathered mushroom data from Europe to create a powerful predictive SDM for North American *Amanita phalloides*. The application of MAXENT software shows promising results for niche modelling based on presence-only data (Wollan et al. 2008) which are often the only data available for fungi. Before applying niche modelling software this study sought to test the quality of DNA data and the available environmental layers.

Studies by Cox et al. (2010a, 2010b) inferred ECM responses to nitrogen deposition at large geographic scales that differ from those at local scales. Here too the argument was made for using DNA to identify ECM in large-scale spatial analysis, but the problems and methodological incongruences of combining multiple studies were also noted. To enable this new facet of mycorrhizal ecology, Lilleskov & Parrent (2007) called for a unified approach to fungal root sampling. We envision that georeferenced fungal DNA sequence data will continue to accumulate rapidly to eventually reveal fungal species distributions. This study explores what signal indicating the environmental preferences of ECM might be already hidden in the growing online databases.

Methods

Twenty different ECM fungi were delimited to species level using ITS DNA sequences from ectomycorrhizas; these were found to be among the most common ECM present at diverse forest and heathland sites (Collier & Bidartondo 2009, Cox et al. 2010a). Location points from Europe were gained using NCBI-BLAST matches from the Genbank and UNITE databases (Fig. 1). High thresholds of similarity were employed (97% for Basidiomycetes and 98% for Ascomycetes - Nilsson et al. 2006) with a minimum sequence coverage of 80% and a minimum sequence length of 400bp to improve confidence of species matches.

In some cases annotated data on GenBank/ UNITE records was used to establish latitude and longitude coordinates, but as this information was often unavailable, associated publications were used to establish source locations. In two cases authors were contacted directly and responded with coordinates (Yarwood SA & Rudawaska M, pers. comm.). Where there were insufficient data for a BLAST match the point was discarded.

Bioclimatic, altitude and soil pH values were extrapolated from the presence of these various fungi using layers obtained from United Nations Spatial Data Infrastructure (nitrogen, soil pH, drainage) WorldClim (bioclimatic and altitude - Hijmans et al. 2005) in ArcGIS. These different species were then tested through randomisation using R version 2.7.2 to examine significant environmental variables. Gathered values were put into a matrix and randomised 1000 times. Where values were below 0.05, the observed environmental variables for that species were considered significantly non-random.

In a separate analysis, presence data gathered for three of the tested fungi were gathered from a source that includes morphologically identified specimen records, the Global Biodiversity Information Facility (GBIF 2009), and compared with the data gathered from DNA databases to address issues of data quantity and quality.
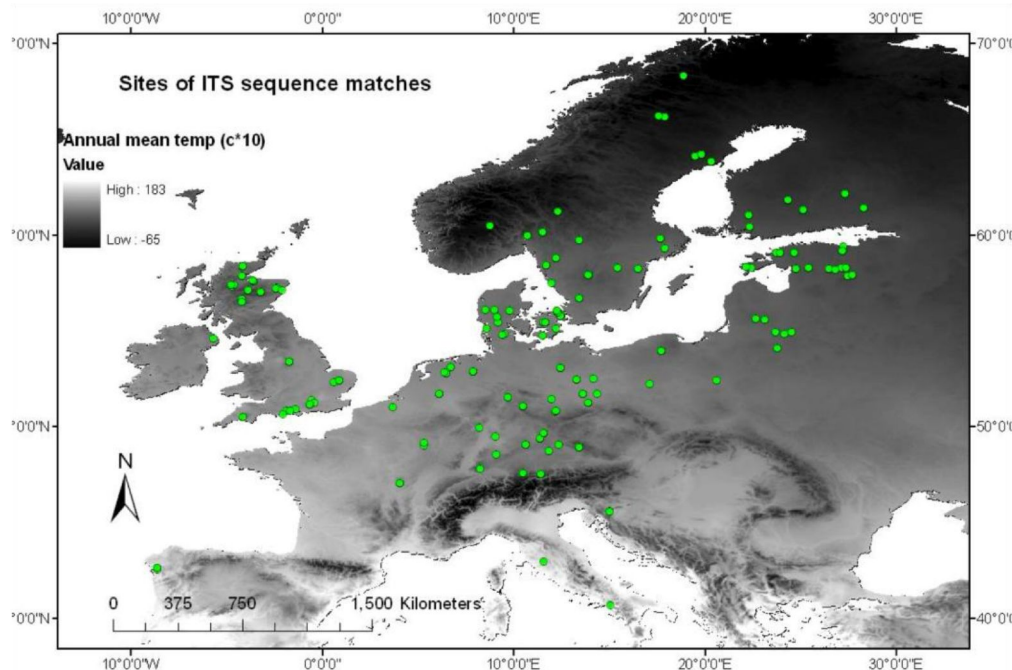
Fig. 1 - Spatial locations of all data gathered from ITS matches using BLAST, from GenBank, UNITE and associated literature.

Results

The total data set for all 20 species was 321 points. Sample sizes per species ranged from 35 *Xerocomus badius* to 9 *Thelephoraceae* spp. The most significant results relate to the annual mean temperature (Tab. 1). There is strong evidence that the results extrapolated from that data are non-random.

When a similar analysis was carried out using data from GBIF, the much larger size of the data set per species should have provided a more representative result (183 for *Elaphomyces granulatus* against 13 BLAST matches). However, these samples suffered heavily from spatially autocorrelated sampling with over half of the samples for *Lactarius rufus* and *Xerocomus badius* originating from Norway. In an attempt to compensate for this, randomised sub-sets of the data were generated and used in the statistical tests.

Discussion

Overall, the bioclimatic variables yielded more significant results than the other environmental layers. Variables such as soil pH have a proven effect on the presence of different ectomycorrhizas (Hung & Trappe 1983), as well as on the presence of different host tree species. The lack of significance when extrapolating from their values in this analysis is likely to be a result of high variability at local scales (*e.g.*, nested pockets of high acidity) and low layer resolution. The significant results found through this analysis, in particular those of the bioclimatic variables (Hijmans et al. 2005), were extrapolated from layers of much higher resolution. These types of variables are more accurately quantified at large spatial scales than soil variables.

Cox et al. (2010b) showed nitrogen as a determinant of fungal diversity across geo-graphical scales but not at a local level. Soil nitrogen was only a significant variable in the present analysis in one case with low sample size; this is most likely also due to layers of low resolution. This can be seen in the generally high levels of variation (Fig. 2). If the annotated information on GenBank records provided information on soil nitrogen, drainage and pH, then the accuracy of environmental layers could be measured by comparing values gained in GIS with those drawn from GenBank and UNITE, prior to statistical testing.

This study sought to test the quality and quantity of data available as much as the data itself; thus, our results show a number of areas which need to be improved for a DNA-based approach to be further used to create SDMs. The extent of this study could soon be improved with new ITS sequence data based on next generation sequencing technologies (Nilsson et al. 2011).

The issue of data quantity and fungal species identification is being addressed through the continual growth of online databases and Hibbett et al. (2011) review the resulting recent progress made in fungal taxonomy. In order to create an SDM a large number of location points is required to verify the respective strengths of environmental variables. That is why this study takes only a preliminary look at a large number of ECM species. Biological GIS data may be subject to three types of bias: taxonomic, temporal and spatial. Gathering data through the use of BLAST aims to reduce taxonomic bias. Although there is variability in the reliability of morphological identification techniques, ITS DNA presents standardized reliable results especially if backed up by multi-locus spe-

Tab. 1 - Fungi from randomisations with significance or near significance for each environmental variable (ITS matches). (~): near significance values of <0.1; (*): randomisation significance of <0.05; (**): randomisation significance of <0.005.

| Variable	Species
pH	*Atheliaceae* sp. (n=22)*
	Thelephora terrestris (n=27)*
	Scleroderma citrinum (n=13)~
Elevation	*Hydnotrya tulasnei* (n=11)*
Drainage	*Cenococcum* sp. (n=14)~
Nitrogen	*Hydnotrya tulasnei* (n=11)*
Annual mean temperature	*Cantharellaceae* sp. (n=18)*
	Elaphomyces granulatus (n=13)*
	Russula paludosa (n=12)*,
	Xerocomus badius (n=34)**
Annual precipitation	*Piloderma* sp. (n=13)*
	Cenococcum sp. (n=14)~
	Elaphomyces granulatus (n=13)~
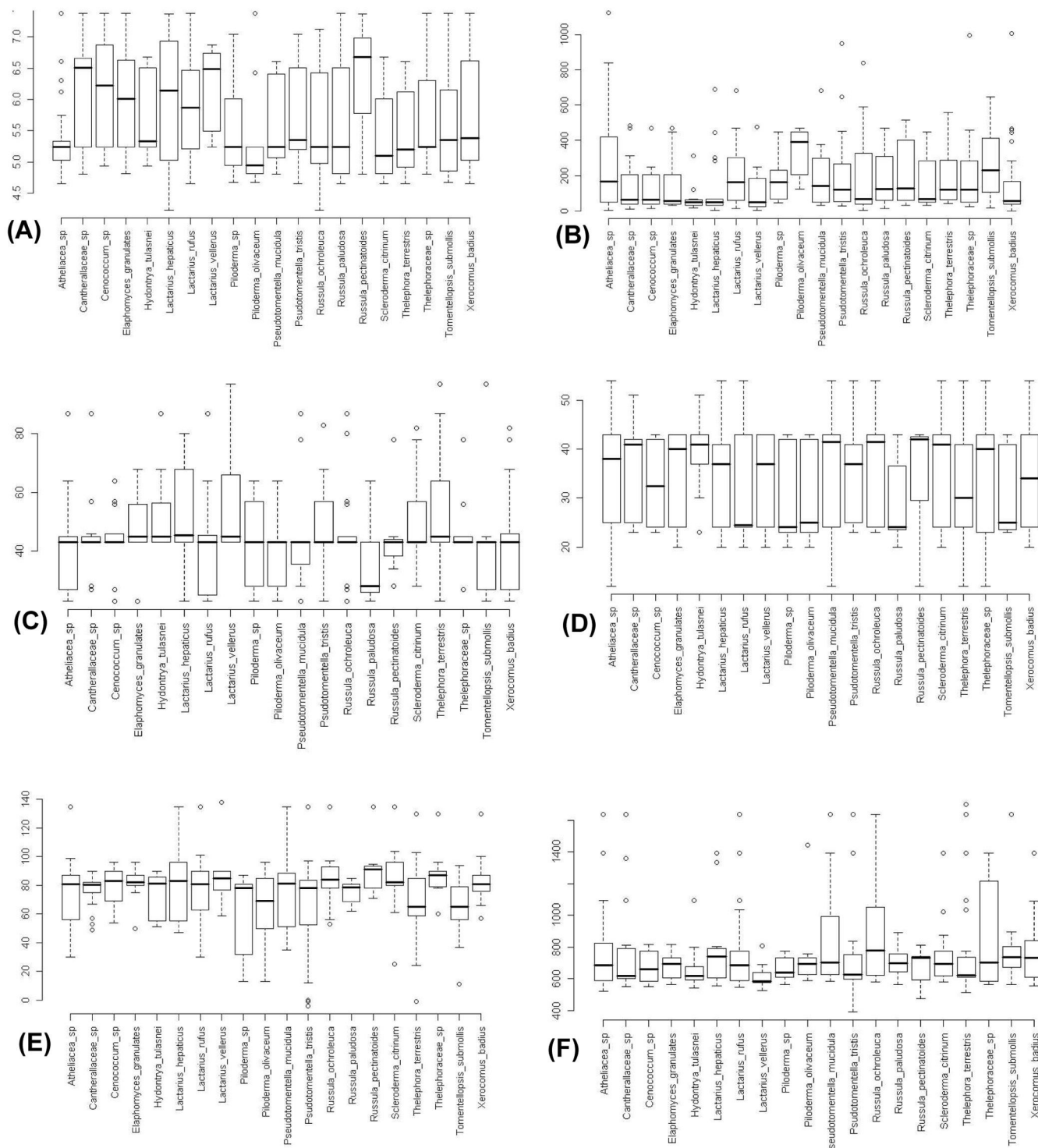
Fig. 2 - Boxplots for environmental data collected from all ITS sequence sites (including F. Cox sites). (A): pH; (B): elevation (meters above sea level); (C): soil drainage (% saturation); (D): soil nitrogen (% of 1%); (E): Annual mean temp (°C x 10); (F): Annual precipitation (mm). Boxes represent inter-quartile range, centre bar represents median, whiskers represent 5th to 95th percentile range.

cies delimitation (*e.g.*, Hedh et al. 2008). Where sporocarp material is relied on, a temporal bias can only be countered by continued sampling effort across fruiting periods. As this is logistically difficult it may be more feasible to use mycorrhizas because they can be temporally stable (Cox 2010, Izzo et al. 2005, Koide et al. 2007). Spatial bias is currently the most detrimental to the use of online databases for creating fungal SDMs and is illustrated by this analysis. The DNA sequences drawn from Genbank and UNITE were predominantly from Denmark, Britain and Sweden even though the original samples were gathered predominantly from

Britain and Germany. Results of significance, extrapolated from fungi *taxa* locations, were not significantly more spatially auto correlated than those *taxa* without significant results. This indicates that spatial bias was not responsible for significant results. Although there are issues of spatial bias inherent in this type of data, they are being addressed through the growth of online databases.

Although the quantity of morphologically-identified data from GBIF was large, and some of the results highly significant, the spatial autocorrelation of the data was also high. Even with a randomised subset of data taken from Norway, the proportion of the data from this area skewed the results. For spatial analysis and particularly for SDMs a large number of locations is required for presence-only data. However, as there was spatial sampling bias for these *taxa*, in addition to the inherent ambiguity of morphologically identified fungal samples, this method would be better served by the growth of fungal databases.

Large scale range maps for ECM only exist for some species at a national level, are based on the presence of fruiting bodies (*e.g.*, Courtecuisse et al. 2008) and are absent from the European Atlas of Soil Biodiversity (Jefferey et al. 2010). A standardised sampling method using DNA identification and gaining data on ECM community composition, soil variables and location would take future analyses closer to SDMs for a multitude of species. Cox et al. (2010a) highlight the potential of ICP Forests for generating uniform data quality. These forests are intensively monitored for biodiversity, atmospheric deposition, soil chemistry, foliar nutrient levels and water balances among other factors across 41 European countries providing both large enough scale and a reliable, scientific resource of historical environmental data for the development of ECM range maps. In addition to this, the data from these sources could be used to create powerful SDMs to predict the presence of ECM species in unsampled areas.

Conclusion

This analysis has shown that the data present in online genetic databases for some ectomycorrhizal fungi can be used to map fungi. However, the validity of this method requires high resolution and accurate environmental layers, an understanding of the variability of environmental factors at different spatial scales and an evenly distributed sampling effort. Low data quantity means

that these results cannot yet be used to make a reliable SDM.

There is need for a standardised level of data collection regarding ECM DNA and the variables of the environment in which they are found. As the strength of spatial data and its extrapolated information is based fundamentally on a larger number of evenly distributed sample locations, the use of online DNA databases provides a reliable means to increase data quality for the development of ECM SDMs.

References

Cox F (2010). The mycorrhizas of Europe's pine forests in the context of nitrogen pollution. PhD thesis, Imperial College London, UK.

Cox F, Barsoum N, Bidartondo MI, Borja I, Lilleskov E, Nilsson L O, Rautio P, Tubby K, Vesterdal L (2010a). A leap forward in geographic scale for forest ectomycorrhizal fungi. Annals of Forest Science 67: 200. - doi: 10.1051/forest/2009107

Cox F, Barsoum N, Lilleskov EA, Bidartondo M (2010b). Nitrogen availability is a primary determinant of conifer mycorrhizas across complex environmental gradients. Ecology Letters 13: 1103-1113. - doi: 10.1111/j.1461-0248.2010.01494.x

Collier F, Bidartondo MI (2009). Waiting for fungi: the ectomycorrhizal invasion of lowland heathlands. Journal of Ecology 97: 950-963. - doi: 10.1111/j.1365-2745.2009.01544.x

Courtecuisse R, Moreau PA, Daillant O (2008). Suivi de la flore fongique: une énorme diversité difficile à mesurer - Partenariat avec les sociétés mycologiques de France. In: "15 Ans de Suivi des Ecosystems Forestiers", Hors Série no. 4, Rendez-Vous Techniques, Office National des Forêts, pp. 99-102.

GBIF (2009). Morphologically identified data for *Lactarius rufus*, *Elaphomyces granulatus* and *Xerocomus badius*. Global Biodiversity Information Facility, Web Site. [online] URL: http://data.gbif.org/datasets

Hedh J, Samson P, Erland S, Tunlid A (2008). Multiple gene genealogies and species recognition in the ectomycorrhizal fungus *Paxillus involutus*. Mycological Research 112 (8): 965-975. - doi: 10.1016/j.mycres.2008.01.026

Hibbett DS, Ohman A, Glotzer D, Nuhn M, Kirk P, Nilsson RH (2011). Progress in molecular and morphological taxon discovery in Fungi and options for formal classification of environmental sequences. Fungal Biology Reviews 25: 38-47. - doi: 10.1016/j.fbr.2011.01.001

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005). Very high resolution interpolated climate surfaces for global land areas. International Journal Climatology 25 (15): 1965 -

1978. - doi: 10.1002/(ISSN)1097-0088

Hung LL, Trappe LM (1983). Growth variation between and within species of ectomycorrhizal fungi in response to pH in vitro. Mycologia 75: 234-241. - doi: 10.2307/3792807

Izzo A, Agbowo J, Bruns TD (2005). Detection of plot-level changes in ectomycorrhizal communities across years in old-growth mixed-conifer forest. New Phytologist 166: 619-629. - doi: 10.1111/j.1469-8137.2005.01354.x

Jefferey S, Gardi C, Jones A, Montanarella L, Marmo L, Miko L, Ritz G, Peres J, Römbke J, van der Putten WH (2010). European atlas of biodiversity. European Commission, Publications Office, Luxembourg.

Koide RT, Shumway DL, Xu B, Sharda JN (2007). On temporal partitioning of a community of ectomycorrhizal fungi. New Phytologist 174: 420-429. - doi: 10.1111/j.1469-8137.2007.02000.x

Lilleskov EA, Parrent JL (2007). Can we develop general predictive models of mycorrhizal fungal community-environment relationships? New Phytologist 174: 250-256. - doi: 10.1111/j.1469-8137.2007.02023.x

Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson KH, Kõljalg U (2006). Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. PLoS ONE 1 (1): e59. - doi: 10.1371/journal.pone.0000059

Nilsson RH, Tedersoo L, Lindahl B D, Kjøller R, Carlsen T, Quince C, Abarenkov K, Pennanen T, Stenlid J, Bruns T, Larsson K-H, Kõljalg U, Kåuserud H (2011). Towards standardization of the description and publication of next-generation sequencing datasets of fungal communities. New Phytologist 191: 314-318. - doi: 10.1111/j.1469-8137.2011.03755.x

Ryberg M, Nilsson R H, Kristiansson E, Topel M, Jacobsson S, Larsson E (2008). Mining metadata from unidentified ITS sequences in GenBank: A case study in *Inocybe* (Basidiomycota). BMC Evolutionary Biology 8 (1): 50. - doi: 10.1186/1471-2148-8-50

Smith SE, Read DJ (2008). Mycorrhizal symbiosis (3rd edn.). Academic Press, London, UK.

Wolfe BE, Richard F, Cross HB, Pringle A (2010). Distribution and abundance of the introduced ectomycorrhizal fungus *Amanita phalloides* in North America. New Phytologist 185: 803-816. - doi: 10.1111/j.1469-8137.2009.03097.x

Wollan AK, Bakkestuen V, Kåuserud H, Gulden G, Halvorsen R (2008). Modelling and predicting fungal distribution patterns using herbarium data. Journal of Biogeography 35 (12): 2298-2310. - doi: 10.1111/j.1365-2699.2008.01965.x