VIEWPOINT

Population proteomics: An emerging discipline to study metapopulation ecology

David G. Biron[1], Hugh D. Loxdale[2], Fleur Ponton[1], Hercules Moura[3], Laurent Marché[4], Christophe Brugidou[5] and Frédéric Thomas[1]

[1] GEMI/UMR CNRS-IRD 2724, Equipe: "Evolution des Systèmes Symbiotiques", IRD, Montpellier, France
[2] Plant and Invertebrate Ecology Division, Rothamsted Research, Harpenden, UK
[3] National Center for Envionmental Health, Centers for Disease Control and Prevention, Division of Laboratory Sciences, Atlanta, GA, USA
[4] INRA, UMR Bio3P, Domaine de la Motte, Le Rheu, France
[5] UMR 5096 (UP-IRD-CNRS), Centre IRD 911, Montpellier, France

Proteomics research has developed until recently in a relative isolation from other fast-moving disciplines such as ecology and evolution. This is unfortunate since applying proteomics to these disciplines has apparently the potential to open new perspectives. The huge majority of species indeed exhibit over their entire geographic range a metapopulation structure, occupying habitats that are fragmented and heterogeneous in space and/or through time. Traditionally, population genetics is the main tool used to studying metatopulations, as it describes the spatial structure of populations and the level of gene flow between them. In this Viewpoint, we present the reasons why we think that proteomics, because of the level of integration it promotes, has the potential to resolve interesting issues specific to metapopulation biology and adaptive processes.

Received: June 7, 2005
Revised: August 2, 2005
Accepted: September 19, 2005

Keywords:
Metapopulation / Population / Population proteomics

Proteomics as a scientific field has, until now, developed in relative isolation from other fast-moving disciplines such as ecology and evolutionary biology. This is unfortunate since applying proteomics to these areas has the potential to reveal new perspectives and lines of research. For instance, one recent review paper shows how proteomics can potentially provide highly valuable information in phylogenetic analyses, permitting the detection and characterization of specific proteins that have evolutionary value in terms of defining mono-, para- and polyphyly [1]. Similarly, some studies have demonstrated how proteomics can be used to investigate natural variations within species populations [2–7].

Correspondence: Dr. David G. Biron, GEMI/UMR CNRS-IRD 2724, IRD, 911 Avenue Agropolis, BP 64501, 34394 Montpellier Cedex 5, France
E-mail: biron@mpl.ird.fr or d.biron@caramail.com
Fax: +33-4-6741-6299

Because of its pivotal role in ecological and evolutionary thinking, adaptation has been extensively studied by all biologists engaged in this area over recorded historical time, indeed from Aristotle onwards. An important aspect of recent theoretical and experimental studies on adaptive processes is the recognition that these phenomena have to be examined in the context of metapopulation structure [8]. The large majority of species exhibit a metapopulation structure over their geographic range, occupying as they do habitats that are fragmented and heterogeneous in space and/or through time. Understanding the ecology of populations at a metapopulation level (*i.e.*, metapopulation ecology) is of interest not only for evolutionary ecologists but also for any scientists who bridge fundamental ecology and applied problem solving, like agronomy and ecosystem management in biological conservation perspectives.

Traditionally, a major goal of population genetics has been, and still is, the measurement of genetic variation within and between populations and the estimation of

fitness differences between genes and genotypes. Population genetics is currently the main approach used to studying metapopulations, since it describes the spatial structure of populations and the level of gene flow between them. A recent nicely designed study in plant genetics suggests the use of 2-DE as a tool to study the adaptive responses to macro-environmental conditions [9]. In the present Viewpoint, we present reasons why we believe that proteomics, especially because of the level of integration it promotes, can be considered as 'molecular phenotyping', and has the potential to resolve interesting issues specific to metapopulation biology and adaptive processes.

To study the proteome of living organisms at the population level, the most popular tool that allows measurement of the genetic variation within and between populations, and hence estimate the fitness difference between genes and genotypes, is 2-DE coupled with MS. In addition, any of the 2-D chromatographic applications considered as second generation of proteomics tools like multi-dimensional protein identification technology (MudPit) and isotope-coded affinity tag (ICAT) can and have been used. Some authors have also referred to 2-DE as an important source of genetic markers [7, 10, 11]. These markers can be quite useful in studies of population genetics, since, as mentioned elsewhere [5–7, 11], the protein spots so resolved are useful monogenic and co-dominant markers that are probably not strongly affected by natural selection. Moreover, 2-DE is a high resolution technique able to separate thousands of genetic products (protein spots) on a single gel and detect isomorphs, polymorphisms and changes such as PTMs (*i.e.*, phosphorylation, glycosylation, acetylation and methylation) induced for instance, by precise ecological situations experienced by individuals [2, 4, 6, 7, 9–11].

Because individuals from different populations of a given species are primarily homogeneous, we expect the huge majority of protein spots resolved by 2-DE to be similar between compared populations. Thus, by definition, population proteomics will be based on the analysis of the genetic polymorphisms detected by a qualitative approach (presence/absence of protein spots) and, secondly, by a semi-quantitative analysis on the abundance (relative volume, vol%) of the protein spots. At the present time, there is a need to develop and test the parameters used to objectively describe population proteomics profiles, along with the necessary statistical tools to quantify their levels of genetic divergence.

For qualitative analysis, a first parameter of interest would be the 'total proteomic richness', defined as the total number of protein spots detected from a representative sample (cell and/or tissue) of individuals belonging to the same species population. As this parameter is highly correlated with the methodological conditions used, it is essential that all populations are analyzed using exactly the same procedure so that comparisons are meaningful. A second parameter would be the specific proteomic richness defined as the number of protein spots specific to a population. Many others parameters of 'classical' population genetics can be used, like Wright's fixation indices: F_{IS} (inter-individual), F_{ST} (subpopulations), F_{IT} (total population) ([5, 7] and M. T. Dorak, http://dorakmt.tripod.com/evolution/popgen.html). It is also possible to test for deviations from Hardy-Weinberg expectations by means of chi-square and exact testing. The exact test is very useful when the sample size and/or some genotype frequencies are small, because it is not based on an asymptotic approximation for large samples [12, 13]. Since it is difficult to homologize loci among treatments using 2-DE (*i.e.*, loci cannot be assumed to be homologous between samples), the genetic distance methods generally employed in population genetic studies cannot be used. A discussion concerning the use of four genetic distance indices (Jacquard, matching coefficient, Nei & Li, corrected matching coefficient) was proposed in a comparison between the A, D, I and R proteomes of the *Triticeae* tribe of cereals [4]. The association coefficient (Nei & Li index) is one of the most used in 2-DE: $F = 2nxy\ (nx + ny)$ where nx and ny are the numbers of protein spots scored in population x and y, respectively, and where nxy is the total number of protein spots shared by both populations x and y. The proteomics distance (D) between populations is $1\text{-}F$ [14, 15]. The use of the Spearman's rank correlation test and Mantel's test (used to correlate between two distance matrices) could then be applied to study the relationships (correlation 'r') between the matrices of proteome distances for populations derived from different geographical areas, and/or to compare the matrix distances obtained using others distances, *e.g.*, geographic, genetic, ecological and/or proteomics distances [16].

For quantitative analysis of protein spots, some statistical descriptive methodologies, such as heuristic analysis, principal component analysis and the Eisen method based on the relative abundance (vol%) of protein spots, can be used to classify the subpopulations and/or populations of a given species in spatial/temporal dimensions [15]. Thereafter, at least two others divergence estimators may be calculated from common protein spots by summing the between-population variances on relative volume (mean variance on abundance) and relative intensity (mean variance on intensity). The mean variance on abundance is the summation of the variances calculated for the relative volume of each common protein spot detected between n populations. The relative intensity of a protein spot is obtained by dividing the relative volume observed for a given common spot by the average of the relative volumes calculated for n common

spots observed for all populations studied. Thus in essence, the mean variance on intensity is the summation of the variances calculated for the relative intensity for each common protein spot detected between *n* populations. The greater the variances are for these estimators, the greater the observed heterogeneities between populations. In addition, the amount of protein variability among populations can be tested using analysis of variance with the relative volume of each protein spot as the dependant variable and the genotype as the factor [9].

Further to these qualitative and quantitative parameters aimed at describing proteomics profiles, special attention should also be devoted to identifying spots, or families of spots, that are characteristic of precise ecological situations experienced by individuals (stress, infection, *etc.*), as well as the resilience of these spots (defined here as the length of time such spots persist in the proteome after their induction). Indeed, among the hundreds of protein spots revealed at any given time on a gel, some have a constant turnover, while others may persist for long periods after their induction, for instance, spots linked to immune processes.

Coupled with so-called 'neutral' DNA markers like microsatellites commonly used in population genetic studies, proteomics markers obtained during population proteomics research could first provide additional information to assess the levels of population structuring (panmixia, discrete subpopulation and isolation by distance) for any given extant species. 2-DE also provides a powerful means to distinguish between very closely related genotypes. Since the protein quantitative variation is often under polygenic control, and subject to epistatic and pleiotropic influences, two genotypes with similar quantitative patterns are unlikely to occur by chance. In comparison to neutral polymorphisms, the predictive power of this kind of polymorphism towards agronomical performances is currently being examined, particularly in plant genetics [3, 9].

Proteomic tools, especially 2-DE, may also identify markers of biological phenomena difficult to study directly in natural populations, but which leave a 'biochemical signature' through time following exposure to biotic (*i.e.*, parasitism and pathogens) and abiotic (*i.e.*, environmental, including chemical, stimuli) selective pressures. For example, population proteomics can be employed to assess contaminant-induced impacts in aquatic systems. Here, many studies performed predict that for any given species, adjacent populations should be more genetically similar than geographically distant ones [17, 18]. Thus, a phenogram of proteomics distances may be useful in delineating water quality groups based on this type of data. The proteomics distance phenogram may thereby depict the differential genetic response populations have to their environment as well as other evolutionary factors influencing genetic differentiation, *e.g.*, restricted gene flow. However, two criteria must be considered to select a species used as bio-indicator. First, species should have a high level of inherent variation. This is because there is a greater probability of identifying appropriate measures for monitoring changes in environmental conditions. Second, species should have broad habitat requirements and be present in sites with a broad range of environmental qualities. If the population distribution of a given species is restricted, then it may be difficult to determine whether environmental quality is a factor limiting this distribution.

Additional research on wildlife epidemiology could greatly benefit from proteomics science. More generally, the identification of stress-specific markers (*e.g.*, famine, drought, and disturbance of one kind or another) has numerous obvious applications in conservation biology. Proteomics tools should also allow enhancement of our understanding of the ecological factors relevant for any given species in an ecosystem/s. Indeed, despite its fundamental importance in ecology, the concept of environmental homogeneity/heterogeneity is difficult to assess, as it typically depends on the species under consideration. Regional environmental variation can result among populations for certain species, and not for others, in different selective regimes and produce 'site-dependent' fitness for phenotype. For a particular species, one way to assess whether two habitats are homogeneous (whatever the level of heterogeneity perceived by the human eye) is to compare the proteome of individuals belonging to these different habitats. The more similar the organisms' proteomes, the more one can safely conclude that individuals are collectively experiencing the same environmental conditions. Fragmented host populations, whereby local selective pressures vary, may prove ideal systems to investigate these phenomena. Figure 1 provides a résumé of the new perspectives apparent for population proteomics.

The apparent separation between sub-disciplines (*i.e.*, genomics and proteomics) that leads to different perspectives on the same ecological reality is a fundamental limitation that needs to be overcome if complex processes, like adaptation, are to be understood. Despite increasing progress towards greater convergence and dialogue between researchers in proteomics and other related disciplines, including population genetics and genomics, much however remains to be done to achieve full integration of these different sub-disciplines when studying adaptive processes. We believe that answers to many current and future questions about population ecology might come as a result of convergence between these disciplines. Population proteomics coupled with population genetics has a great potential to resolve issues specific to the ecology and evolution of natural populations. Presently, this promising area of research is in its infancy and clearly, a much larger global research effort is required to 'calibrate' the methodology and concepts. We need to understand the meaning and the reliability of the different kinds of proteomics markers employed for studying metapopulations. There are undoubtedly many direct and/or indirect correlations between the expression of proteomic spots and a large range of biological processes that are of interest in population ecology. Even if we are as yet far from this 'promised land', a better understanding of the informa-

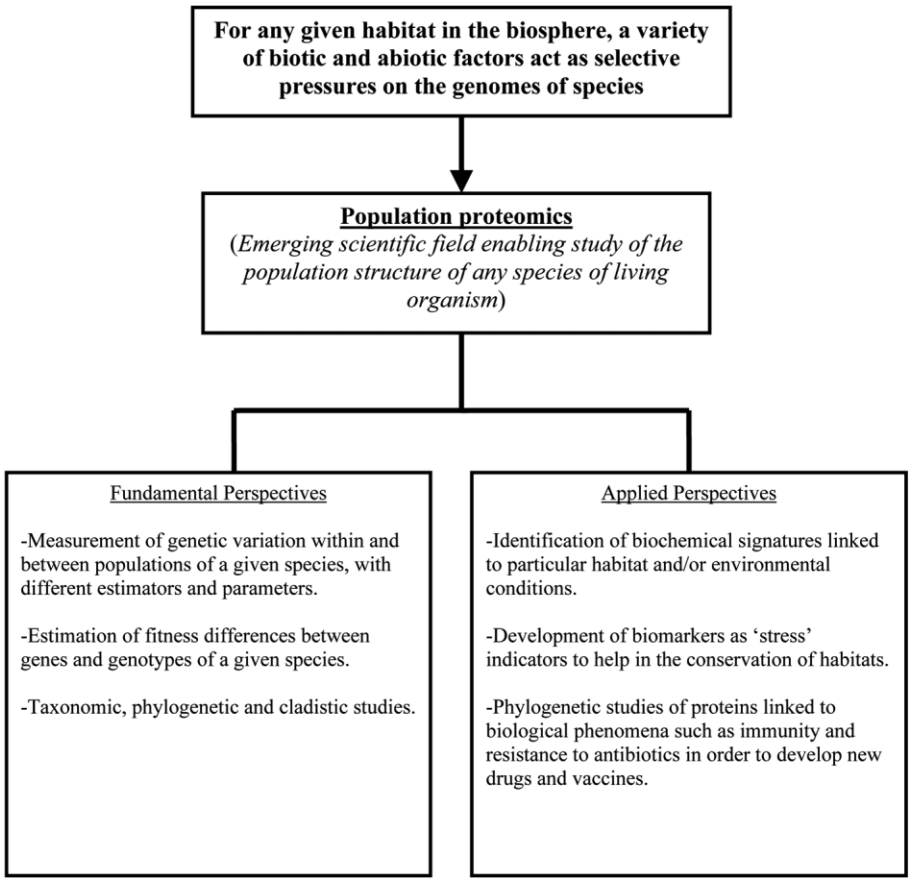

Figure 1. Potential of population proteomics as an emerging discipline.

tion contained in proteomics markers should permit an impressive amount of information to be gathered on the past as well as current ecological conditions experienced by a given population, something that could be summarized as 'show me your proteome and I will tell you who you are, where you are from, and where you should go from here'.

References

[1] Navas, A., Albar, J. P., *Proteomics* 2004, *4*, 299–302.

[2] Bahrman, N., Zivy, M., Damerval, C., Baradat, P., *Theor. Appl. Genet.* 1994, *88*, 407–411.

[3] Burstin, J., De Vienne, D., Dubreuil, P., Damerval, C., *Theor. Appl. Genet.* 1994, *89*, 943–950.

[4] Zivy, M., Madidi, S. E., Thiellement, H., *Electrophoresis* 1995, *16*, 1296–1300.

[5] Mosquera, E., López, J. L., Alvarez, G., *Heredity* 2003, *24*, 432–442.

[6] Chevalier, F., Martin, O., Rodifal, V., Devauchelle, A. D. *et al.*, *Proteomics* 2004 *4*, 1372–1381.

[7] López, J. L., *J. Chromatogr. B.* 2005, *815*, 261–274.

[8] Hanski, I., *Ecology and Evolution*, Oxford University Press, Oxford 2002.

[9] David, J. L., Zivy, M., Cardin, M. L., Brabant, P., *Theor. Appl. Genet.* 1997, *95*, 932–941.

[10] Leigh Brown, A. J., Langley, C. H., *Proc. Natl. Acad. Sci. USA* 1979, *76*, 2381–2384.

[11] Touzet, P., Morin, C., Damerval, C., Le Guilloux, M. *et al.*, *Electrophoresis* 1995, *16*, 1289–1294.

[12] Louis, E. J., Dempster, E. R., *Biometrics* 1987, *43*, 805–811.

[13] Guo, S. W., Thompson, E. A., *Biometrics* 1992, *48*, 361–372.

[14] Tastet, C., Bossis, M., Renault, L., Mugniéry, D., *Nematology* 2000, *2*, 343–353.

[15] Biron, D. G., Moura, H., Marché, L., Hughes, A. L., Thomas, F., *Trends Parasitol.* 2005, *21*, 162–168.

[16] Fullaondo, A., Vicario, A., Aguirre, A., Barrena, I., Salazar, A., *Heredity* 2001, *87*, 266–272.

[17] Mayr, E., *Population, Species and Evolution*, Havard University Press, Cambridge 1970.

[18] McArthur, J. V., Leff, L. G., Smith, M. H., *J. N. Am. Benthol. Soc.* 1992, *11*, 269–277.

