

# Next-Generation Genome Sequencing Makes Non-Model Organisms Increasingly Accessible for Proteomic Studies: Some Implications for Ecotoxicology

David Sheehan\*

Proteomics Research Group, Department of Biochemistry, University College Cork, Ireland

A key step in proteomics is matching of sequences obtained from tandem mass spectrometry of (usually) tryptic peptides to genomes held in public domain databases [1]. Such genomes first became available from important model organisms such as human, mouse, rat and a range of other representative animal, plant and microbial species [2]. These model species therefore became accessible for proteomic studies and rich and reliable data could be obtained. However, scientists working with non-model organisms often failed to identify proteins reliably notwithstanding the availability of *de novo* sequence data from tandem MS because non-model species tend to be poorly-represented in public-domain sequence databases [3]. Often, such studies have had to rely on sequence similarity with organisms represented in sequence databases rather than sequence identity and sometimes, no reliable match was possible even with this approach. However, recent improvements in high-throughput genome sequencing such as “next-generation sequencing” have significantly increased the rate of genome sequencing which now outpaces Moore’s Law, more than doubling each year [4]. For example, in 2007, a single sequencing run could generate 1 Gbp of sequence data but, by 2011, such a run could generate a Tbp of data, almost one thousand times more in just four years. This has greatly increased the speed of genome sequencing and the list of whole genomes grows apace [2].

A good example of use of non-model organisms is in environmental toxicology (ecotoxicology) where “sentinel species” have long been used to assess the pollution status of natural environments [5]. The idea is that such organisms may show population-level, physiological or biochemical effects when stressed by adverse environmental conditions due to climate change or exposure to chemical pollutants. Much interest has centred on aquatic organisms inhabiting freshwater, estuarine or coastal habitats since a majority of human settlement worldwide tends to be within a relatively short distance of coastal zones. Thus, key sources of pollution and environmental degradation such as industrial, agricultural and shipping activities are often adjacent to rivers, estuaries and coastlines. More recently, there is growing concern about the implications of climate change, habitat loss, deterioration of water quality and ocean acidification arising from anthropogenic (man-made) pollutants, especially emerging categories such as endocrine disruptors [5], nanomaterials [6], and human/animal pharmaceuticals [7]. There was a tendency to use robust organisms capable of withstanding a range of pollution and environmental stress but with a widespread geographical distribution worldwide and relatively sessile life-style. Examples would include bivalves such as the genus *Mytilus* (in estuaries) and the water-flea *Daphnia* (freshwater). Since the 1960’s much work has concentrated on identification of biochemical indices of damage (biomarkers) within such species such as lipid peroxidation or changes in stress-response protein profiles. In principle, proteomics provides a welcome complement to these classical approaches [8,9] but there is a problem—many aquatic species have especially large genomes

(as big as or even bigger than the human genome). Thus, they were poorly-represented in public-domain sequence databases and this severely limited the potential of proteomics approaches in ecotoxicology [8,9]. However, the accelerating pace of genome sequencing described above has recently made available the full genomes of *Daphnia pulex* [10] and the pacific oyster, *Crassostrea gigas* [11], while that of *Mytilus* is eagerly awaited. These developments have significantly improved prospects for proteomics as a tool in ecotoxicology research into the future. This will make possible identification of novel biomarkers and elucidation of toxicity mechanisms as well as data-rich explorations of effects on entire protein networks.

## References

1. Nesvizhskii AI, Vitek O, Aebersold R (2007) Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods* 4: 787-797.
2. <http://www.ncbi.nlm.nih.gov/genome>
3. Dowd WW (2012) Challenges for biological interpretation of environmental proteomics data in non-model organisms. *Integr Comp Biol* 52: 705-720.
4. Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26: 1135-1145.
5. Porte C, Janer G, Lorusso LC, Ortiz-Zarragoitia M, Cajaraville MP, et al. (2006) Endocrine disruptors in marine organisms: Approaches and perspectives. *Comp Biochem Physiol C Toxicol Pharmacol* 143: 303-315.
6. Tedesco S, Sheehan D (2010) Nanomaterials as emerging environmental threats. *Curr Chem Biol* 4: 151-160.
7. Rosi-Marshall EJ, Royer TV (2012) Pharmaceutical compounds and ecosystem function: An emerging research challenge for aquatic ecologists. *Ecosystems* 15: 867-880.
8. Dowling VA, Sheehan D (2006) Proteomics as a route to identification of toxicity targets in environmental toxicology. *Proteomics* 6: 5597-5604.
9. Lemos MFL, Soares AMVM, Correia AC, Esteves AC (2010) Proteins in ecotoxicology – How, why and why not? *Proteomics* 10: 873-887.
10. Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, et al. (2011) The ecoresponsive genome of *Daphnia pulex*. *Science* 331: 555-561.
11. Zhang G, Fang X, Guo X, Li L, Luo R, et al. (2012) The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490: 49-54.

\*Corresponding author: David Sheehan, Proteomics Research Group, Department of Biochemistry, University College Cork, Western Gateway Building, Western Rd. Cork, Ireland, Tel: +353 21 4205424; E-mail: [d.sheehan@ucc.ie](mailto:d.sheehan@ucc.ie)

Received January 02, 2013; Accepted January 04, 2013; Published January 07, 2013

Citation: Sheehan D (2013) Next-Generation Genome Sequencing Makes Non-Model Organisms Increasingly Accessible for Proteomic Studies: Some Implications for Ecotoxicology. *J Proteomics Bioinform* 6: e21. doi:10.4172/jpb.10000e21

Copyright: © 2013 Sheehan D. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.