Indian J. Plant Genet. Resour. 35(3): 289–292 (2022)
DOI 10.5958/0976-1926.2022.00084.5

Informatics as Future for Fish Genetic Resources Management

Simon Wilkinson
Network of Aquaculture Centres in Asia-Pacific, Bangkok, Thailand

Informatics is the application of data science to extract and communicate actionable knowledge in the context of a specialised domain. Informatics will play an increasingly important role in management of genetic resources for aquaculture as the industry pursues automation and optimisation of processes in pursuit of intensification. This will likely parallel the adoption of industrial control systems in technology-oriented "smart farms", which will drive instrumentation of farming systems. Aquaculture has historically under-invested in genetic resource management and all aspects of information technology, including informatics, but both have potential to generate substantial benefits. Data capture and computational resources for informatics are no longer limiting. They are cheap, easy to obtain and manage, due to advances in microelectronics, software virtualisation and cloud computing. Human resources are now the main bottleneck in application of informatics in aquatic genetic resources management. There is a need for the aquaculture sector to invest in developing its own professional capacity in informatics, as it is unlikely to be able to recruit personnel from more industrialised, higher-paying sectors

Introduction

Informatics has its foundations in data science, collecting, organising, curating and analysing information to extract actionable knowledge that can be used to inform decision making and address specific problems (Fridsma 2018). Informatics is also concerned with the design of information systems and tools to interpret, visualise and communicate knowledge from data. Informatics is distinguished from 'pure' data science by the requirement to have a degree of specialist knowledge in a domain to understand the issues relevant to that domain and to devise appropriate collection, analyses and applications. Hence there are many specialisations within informatics, such as veterinary informatics, breeding informatics and so on.

We presently stand at the beginning of a revolution in computational informatics. This revolution is being driven by colossal leaps in automated data capture and storage; the availability of massive and affordable cloud-based computational processing on demand; and near-ubiquitous electronic network communications. Collectively, these allow informatics to operate on a scale that was unimaginable, just a few years ago. We have never had so much information. Our problem, now, lies in how to generate knowledge from it. This in turn requires public and private investment in both people and computing infrastructure.

Informatic Opportunities in Fish Genetic Resources Management

Informatics can play a valuable role in nearly every aspect of aquaculture and will increasingly do so as the industry pursues automation and optimisation of processes in pursuit of intensification. In the context of genetic resources management (setting aside farm management), informatics can play a role in many ways.

In aquaculture, non-exhaustive examples include:

- Analysis of molecular genetic sequences, markers, genotypes and phenotypes, production traits, pedigrees and similar information in broodstock holdings that is of interest to farm productivity and/or conservation genetics. In aquaculture, such information can inform selective breeding programmes that improve growth rates or favourable production characteristics of farmed animals by selecting appropriate breeders.
- Quantifying the performance of animals with different genotypes, phenotypes and traits, in the context of particular farm conditions or management regimes. For example, growth and survival at different temperatures, salinities, resistance to specific diseases or performance on certain nutritional regimens.

*Author for Correspondence: Email-simon@enaca.org

With regards to management of wild genetic resources, recording and curating data such as:

- Fisheries catch and effort.
- Distribution and habitat records (particularly for threatened species).
- Environmental monitoring and management data.
- General taxonomy and biology data.

These are a range of very different problems with very different sources of data that are likely to be carried out by informatics specialists with a diverse range of experience, many of whom will use other professional titles.

One area of emerging interest is the application of on-farm informatics. This will likely parallel the adoption of industrial control systems in technology-oriented "smart farms". While it already occurs to a degree through assessment of common farm parameters such as survival, feed conversion ratio and similar, advances in microelectronics, network communications and sensor technologies will increasingly drive the use of automation and industrial control systems in farm management. The instrumentation associated with these developments will permit real time monitoring and capture of detailed data from the farm environment, facilitating more complex analyses and optimization of farming systems and practices, and their interaction with genetic variation of stock.

Aquatic Genetic Resources Lagging Behind

As a relatively new industry, aquaculture lags well behind terrestrial agriculture in most respects, such as in the adoption of traditional/standard practices such as selective breeding and in more recent innovations such as automation.

In terrestrial agriculture, selective breeding is taken for granted. Everything you eat is an improved variety. Science-based selection has been underway for many decades, and informal selection for desirable characteristics has been underway for centuries or millennia.

More recently, in some industries, such as dairy cattle, commercial genomics services are available to evaluate individual breeders or even embryos against desirable traits including yield, fertility, survival, heat tolerance, feeding efficiency and others, enabling farmers to select high quality genetic material that best suits their business needs and priorities.

By contrast, aquaculture has under-invested in genetic resource management, and very few improved varieties are available in aquaculture. Genetic mismanagement and inbreeding are rampant. In many cases we are farming degraded genetic resources that have lower performance than unimproved, wild-type animals. Wild-sourced seed are still commonly used in some sectors of aquaculture.

Aquaculture has also under-invested in all aspects of information technology, including informatics. The reasons for this are not clear, but as an emerging agricultural sector it is perhaps logical that most effort should be put into solving the biological problems first. Secondly, with little to no automation and the small-scale nature of most farms, the field has attracted few information technology specialists, and is dominated by biology graduates.

But there is some good news. Advances in IT are making informatics more accessible

In the recent past, setting up infrastructure for computational informatics required the purchase or lease of a server, physical or virtual, its co-location in a commercial data centre or on premises with an expensive lease line connection to the internet, and you were limited to whatever hardware you could afford at the time. Data acquisition and entry, particularly from field sources, was largely a manual process.

Recent advances in microelectronics, sensor technologies, network communications and software virtualisation, containerisation and orchestration have radically improved the accessibility of computational informatics by slashing the cost and maintenance requirements, while making the hardware available to anyone via cloud computing.

Networked Sensors and Microcontrollers

From an informatics point of view, the availability of cheap sensors and microcontrollers has opened the possibility of adding industrial control systems to aquaculture production systems. This in turn allows for 24/7 automated capture of data for real-time monitoring or historical analysis of trends.

Common sensors such as temperature, pH, ammonium, turbidity, motion sensors and many more are typically a few dollars (dissolved oxygen is an exception) and can be connected to programmable microcontrollers, which may relay data to a networked collection point, exercise control over hardware such as pumps, motors and aerators for farm management, or trigger warnings if environmental parameters cross acceptable thresholds.

The recently released Raspberry Pi Pico W microcontroller, for example, costs only $6, includes WIFI connectivity for easy networking, is programmable in Python or C and has excellent documentation and a community of makers and users from whom advice can be sought.

Cloud-based Server Infrastructure

From a computational informatics perspective, it is no longer necessary to own or maintain server hardware. Cloud-based computing providers allow you to set up, tear down, and change the specifications of servers at a moment's notice.

Opening an account with a cloud-based computing provider just takes a couple of minutes and a credit card. Having done that, you are free to commission *as many servers as you need*, on demand, whether virtualised or physical. You can choose the location of servers from a list of data centres around the globe. You can specify what operating systems you want them to run, the number of CPUs and how much RAM you need for the task at hand. Dedicated GPUs are available to support applications that require high levels of parallel processing.

From specifying requirements to provisioning a new server to spinning it up with a first-class internet connection takes a few minutes. You will typically be billed by the hour at a rate that is astonishingly cheap and would be impossible to achieve purchasing your own equipment, from as little as $5 per month. Once your task is accomplished, you can decommission the servers and walk away without paying anything more. Or come back later and re-commission servers when you have need. Automatic provisioning and scaling of servers in response to load is possible with many providers.

These are the basics, most cloud-based computing providers offer a full range of services that are beyond the scope of this discussion, but for practical purposes if you are likely to need it, they are likely to have it.

Containerisation

Server resources can be efficiently partitioned still further through containerisation technologies, such as Docker, which enable the creation of very light-weight virtual machines or "containers". The mechanism is beyond the scope of this paper, but essentially a Docker container makes use of the Linux kernel already running on the host machine, rather than requiring a full copy of all the files and a separate running instance, so it is extraordinarily efficient. A container's processes are assigned to a separate namespace, so they are logically separated from the host and operate as an independent, virtualised machine.

The strength of Docker and a related tool, Compose, is that an entire network of servers and services can be built from a configuration file, in which you specify the software, data volumes, and network topology you require. Docker and Compose will retrieve the relevant server images, set up containers, virtual network connections and persistent data storage volumes with a single command. Since data volumes are persistent, containers (servers) can be shut down, discarded, rebuilt or swapped out as needed. The container infrastructure is essentially disposable and readily upgradeable, simplifying maintenance greatly.

Orchestration

Having established a fleet of servers and / or containers, maintaining them and enforcing standardised configurations could be difficult. But in a manner similar to Docker as described above, free software tools such as Ansible can be employed to automate and standardise deployments, reducing the task to preparation of a common specification file ('playbook' in Ansible terminology), which will implement a prescribed series of actions against designated groups of servers and report their status and any variation from the prescribed configuration.

Such software enables a single administrator to manage potentially hundreds of machines, a task that used to take a significant work force.

We still have a people problem

The above discussion above boils down to this: Data capture and computational resources for informatics are no longer limiting. They are cheap, astonishingly easy to obtain, and much easier to maintain than ever. So why isn't informatics being broadly applied in aquaculture?

The slow uptake of informatics in aquaculture generally, including for management of genetic resources, is primarily due to a lack of personnel with relevant training and expertise. The availability of computer hardware, software and networks has outstripped the available human resource capacity; it is the people with the knowledge and skills to harness these tools that is lacking.

It is a fact that aquaculture is, relatively speaking, a low paid industry. Most commercial farms simply lack the scale and infrastructure to require and support advanced information technology professionals. Graduates in informatics, particularly those with a genetics background, will certainly look to more technologically advanced industries and the more highly paid jobs available in larger companies for their careers, for the foreseeable future.

We Must Look Forward

What then, to do? In the authors opinion, the foundations of informatics should be taught as a component of aquaculture at tertiary level institutions, building on existing record keeping components. Aquaculture and conservation-oriented research schools should consider establishing related streams in informatics to support the evolution of aquaculture to a "smart", technologically based industry, with a view to expanding informatics offerings over time as industry demand grows. It will take time, but there will never be a better time to start than right now.

It is likely that aquaculture will benefit from spill-over of innovation from other sectors and may itself become a source of expertise for aquatic conservation. But to lead the way in solving aquaculture and aquatic conservation problems it would be better for the industry to develop its own native capacity in informatics.

References

Besson M, H Komen, G Rose and M Vandeputte (2020) The genetic correlation between feed conversion rate and growth rate affects the design of a breeding program for more sustainable fish production. *Genet. Sel. Evol.* **52**(5): DOI: https://doi.org/10.1186/s12711-020-0524-0

Fridsma DB (2018) Data sciences and informatics: what's in a name? *J. Am. Med. Inform. Assoc.* **25**(1): 109. doi: 10.1093/jamia/ocx142

Smith RD and M Williams (2000) Applications of informatics in veterinary medicine. *Bulletin Medical Library Association* **88**(1): 49-51.