Langer *et al. Genome Biology* (2025) 26:228
https://doi.org/10.1186/s13059-025-03673-9

Genome Biology

REVIEW **Open Access**

Empowering bioinformatics communities with Nextflow and nf-core

Björn E. Langer[1], Andreia Amaral[2,3], Marie-Odile Baudement[4], Franziska Bonath[5], Mathieu Charles[6,7], Praveen Krishna Chitneedi[8], Emily L. Clark[9], Paolo Di Tommaso[10], Sarah Djebali[11], Philip A. Ewels[10], Sonia Eynard[12], James A. Fellows Yates[13,14,15], Daniel Fischer[16], Evan W. Floden[10], Sylvain Foissac[12], Gisela Gabernet[17], Maxime U. Garcia[10], Gareth Gillard[4], Manu Kumar Gundappa[9], Cervin Guyomar[12], Christopher Hakkaart[10], Friederike Hanssen[18,19,20], Peter W. Harrison[21], Matthias Hörtenhuber[22,23], Cyril Kurylo[12], Christa Kühn[8,24], Sandrine Lagarrigue[25], Delphine Lallias[6], Daniel J. Macqueen[9], Edmund Miller[26], Júlia Mir-Pedrol[19], Gabriel Costa Monteiro Moreira[27], Sven Nahnsen[18,19,20,28], Harshil Patel[10], Alexander Peltzer[29], Frederique Pitel[12], Yuliaxis Ramayo-Caldas[30], Marcel da Câmara Ribeiro-Dantas[10,31], Dominique Rocha[6], Mazdak Salavati[9,32], Alexey Sokolov[21], Jose Espinosa-Carrasco[1*], Cedric Notredame[1,33*] and the nf-core community

*Correspondence:
joseantonio.espinosa@crg.eu;
cedric.notredame@crg.eu

[1] Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain
Full list of author information is available at the end of the article

Abstract

Standardized analysis pipelines contribute to making data bioinformatics research compliant with the paradigm of Findability, Accessibility, Interoperability, and Reusability (FAIR), and facilitate collaboration. Nextflow and Snakemake, two popular command-line solutions, are increasingly adopted by users, complementing GUI-based platforms such as Galaxy. We report recent developments of the nf-core framework with the new Nextflow Domain-Specific Language (DSL2). An extensive library of modules and subworkflows enables research communities to adopt common standards progressively, as resources and needs allow. We present an overview of some of the research communities built around nf-core and showcase its adoption by six EuroFAANG farmed animal research consortia.

Introduction

Advancements in large-scale molecular biology methods have driven an unprecedented increase in data generation [1–3]. This situation brings computational capacities to their limits and highlights a growing need for robust and scalable data analyses [4]. Workflow management systems (WfMSs) are now the recommended solution when dealing with high-throughput data analysis pipelines [5]. While some frameworks, such as Galaxy [6], focus on user-friendly graphical interfaces, other tools like Snakemake [7] and Nextflow [8] are designed for bioinformaticians familiar with programming. They combine the expressiveness [9] of Bash with additional features to better support reproducibility, traceability, parallelization, and portability across different infrastructures (HPC clusters, cloud computing, and workstations). Using Google Scholar to explore the bibliometrics of some of the most popular WfMSs [9], we measured 4706 citations for the year 2024, which is about two and a half times more than for 2018 (Fig. 1; Additional File 1: Supp. Table 1). A notable trend over the last 6 years has been the significant increase in Nextflow and Snakemake usage, while Galaxy has remained relatively stable in absolute citation numbers after peaking in 2021. These trends are corroborated by a similar analysis using Scopus data (Additional File 1: Supp. Figure 1; Additional File 1: Supp. Table 2) and align with the findings from the *2024 State of the Workflow* survey conducted by Seqera (Additional File 1: Supp. Table 4). This survey, involving 608 Nextflow users across 48 countries, revealed a notable decline in the proportion of respondents declaring some usage of Galaxy—from 17% in 2021 to 5% in 2024. During the same period, the proportion of Snakemake users decreased from 27 to 17%. Our bibliometric analysis also indicates that Nextflow experienced the highest growth in usage among WfMSs. With a citation share of about 43% in 2024, Nextflow has become the main driver behind the adoption of bioinformatics-based WfMSs. All the WfMSs mentioned here are open-source software available on GitHub, where repositories can be evaluated by users through a star system (GitHub equivalent of a "like"). Based on their number of stars, the WfMSs surveyed here rank among the top 30,000 repositories (out of an estimated 420 million [10]). Notably, Nextflow is the highest-ranked among them, positioned at #3210 with 4032 stars (Additional File 1: Supp. Table 3). The status of Nextflow as a major WfMS is also reflected in global registries such as the WorkflowHub [11], an ELIXIR-supported initiative [12] aimed at maintaining a registry for scientific computational workflows (Additional File 1: Supp. Table 3). In 2024, Nextflow pipelines accounted for 24.1% of WorkflowHub entries and were the second most

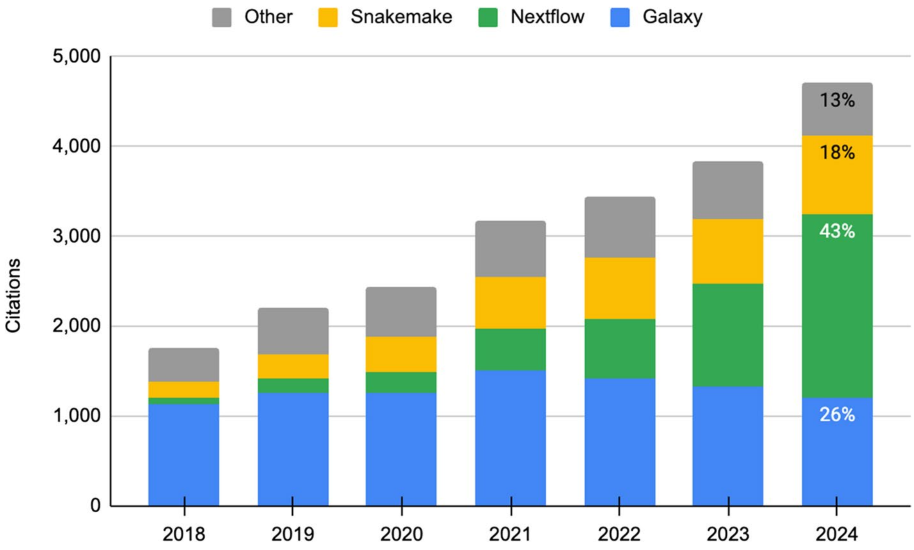

Fig. 1 Google Scholar citation counts for bioinformatics workflow management systems. Sum of citations of the major publications of Galaxy, Nextflow, and Snakemake between 2018 and 2024 (Data in Additional File 1: Supp. Table 1)

represented after Galaxy, which accounted for 50.8%. Altogether, these observations underscore the global relevance of WfMSs to the community and highlight the central role of Nextflow.

While WfMSs provide a powerful way to bundle available methods into pipelines, they do not provide standards for how this bundling should be carried out. This gap has led to the establishment of pipeline registries for Galaxy (Galaxy ToolShed [13]), Snakemake (Snakemake Workflow Catalog (SWC) [14]; Snakemake-Workflows [15]), and Nextflow (nf-core [16]). These workflow repositories contain pipelines developed by the community and made available to everyone, with a focus on usability and long-term maintenance. Registries promote consistent pipeline design by defining clear best-practice guidelines (Additional File 1: Supp. Table 3). "Automated reproduction"—the capacity of programmatically deploying a pipeline off the shelf without a crash—was recently quantified by an independent study comparing nf-core and Snakemake SWC [17]. The authors reported that 83% of nf-core's released pipelines could be deployed as expected, a figure nearly four times higher than that reported for SWC.

Altogether, this data suggests a rapidly changing landscape for the use of WfMSs. Biologists are increasingly confronted with the complex task of selecting a technology with which they can be confident their data will be exploited at its full potential. They must also ensure that their results will remain interoperable over extended periods. The rapid pace of data and methods evolution also means that many stakeholders will alternate between being users and developers of existing pipelines, if only to adapt existing tools to their needs. It is, therefore, essential that the decision to commit to a WfMS is as well-informed as possible. In the next section, we outline recent developments in Nextflow and nf-core and explore aspects relevant to both users and developers. In the last section, dedicated to communities, we explore the decision-making process and the possibility of undergoing what we refer to as *progressive standard transition*, a process we expect to be relevant to any community, especially those with limited resources.

Nextflow and nf-core: recent developments

Nextflow and nf-core are community resources collectively maintained by academia and industry. In this section, we review the main developments that took place between 2018 and 2025.

Expansion of the nf-core pipeline community

The nf-core community was created in 2018 for the specific purpose of maintaining a curated collection of pipelines implemented according to agreed-upon best-practice standards [18]. The nf-core pipelines are meant to be off-the-shelf tools and are characterized by reproducibility, standardization, and rapid result generation. To achieve this, the community selects widely used state-of-the-art bioinformatics tools. These tools are subject to peer review and community feedback and are further developed under a well-defined decision-making process (see the next section). As of February 2025, there are 124 pipelines available within nf-core, covering the analysis of a broad range of data types, including high-throughput sequencing, mass spectrometry, protein structure prediction, or microscopy, but also applications well beyond biology such as astrophysics [19], earth science [20], and economics [21]. These curated pipelines are supported by over 2600 GitHub contributors (including > 1200 organization members) and over 10,000 users on nf-core's primary communication platform Slack (Additional File 1: Supp. Table 6). Pipeline long-term maintenance is achieved by an efficient cooperation model allowing teams across institutions to collectively and simultaneously contribute according to their needs and capacity (Fig. 2A).

nf-core governance and broadening of its outreach

The nf-core community is organized around a steering committee (currently 7 members, including 3 Seqera employees) that provides guidance and recommendations and a core team (currently 14 members, including 7 Seqera employees) that ensures the day-to-day running of the nf-core project, oversees governance teams, and makes routine

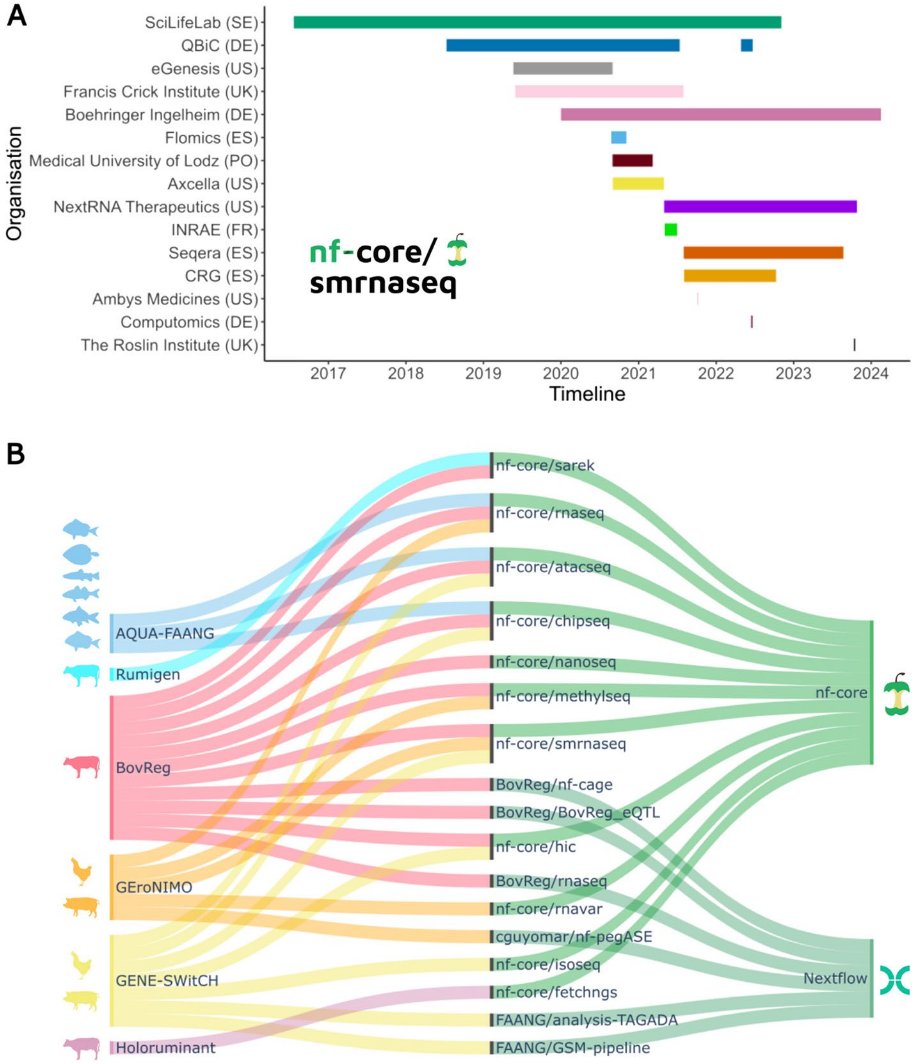

Fig. 2 Pipeline maintenance and usage. **A** Major contributions to the nf-core/smrnaseq pipeline over time by different academic institutions or private companies. Data for individual contributors is collapsed to their institution (SciLifeLab: 3; QBiC: 2; Boehringer Ingelheim: 3; Seqera: 4; all the others: 1) **B** Nextflow analysis pipelines used in the EuroFAANG consortia for the functional annotation of various species' genomes

community decisions (Additional File 1: Supp. Table 4). The community features contributors from both academia and industry and benefits from a strong material support from Seqera that provides key infrastructure elements, such as the Slack channel or the organization of annual gatherings. Together, the core team organizes every aspect of the community, including the incorporation of teams dedicated to overseeing infrastructure maintenance, safety, and outreach. Decision-making is generally based on community discussion via Slack. For instance, new pipeline projects or suggestions for modifications to existing ones are first discussed on Slack and community-approved. These suggestions are then implemented and proposed via GitHub pull requests (e.g., new content proposals for a pipeline repository). For acceptance, these pull requests must be reviewed and approved by at least two nf-core members in the case of a release pull request and at least one otherwise. The core members also manage outreach efforts by organizing the compilation of community-wide resources through a dedicated team. For instance, free community Nextflow and nf-core training videos are recorded and shared online every 6 months, accounting for over 35,000 views on YouTube between 2022 and 2023 (Additional File 1: Supp. Table 5). Inclusiveness is a central aspect of nf-core as reflected by the organization of events accessible to all, irrespective of language, geography, time zones, and the ability to travel. Since its inception, nf-core has run 16 virtual, hybrid, or in-person hackathon events that have attracted global attendance. Over 100 "nf-core/bytesize" webinars describing community advancements, coding approaches, and pipelines have been shared online. A mentorship program pairs experienced Nextflow and nf-core developers with new community members from underrepresented groups. The effectiveness of this effort is reflected in the various metrics collected across resources. Overall, views of outreach material on YouTube increased from 6761 in 2022 to 28,398 in 2023 (Additional File 1: Supp Table 5). This evolution is on par with the near-doubling of GitHub contributors (1156 in 2022 up to 2602 in 2025) and the number of registered Slack users that has increased more than tenfold (782 vs 10,856) since the July 2020 release of the new Nextflow syntax (Domain-Specific Language 2, DSL2) (Additional File 1: Supp. Table 6; Additional File 1: Supp. Section DSL2).

Evolution of the Nextflow workflow management system

DSL2 has been Nextflow's most notable technical development over the past 6 years (Additional File 1: Supp. Section "DSL2" for a technical description). From a developer's point of view, DSL2 supports the same functionalities as DSL1, with its key improvement being to allow the splitting of complex workflows into smaller modular components. This includes modules (encapsulating a specific computational task) and subworkflows (orchestrated groups of module tasks), which are both reusable across multiple workflows. Other main developments include improved support for cloud-based computing platforms, support for on-demand container provisioning via the Wave service [22], and extended support for additional HPC schedulers/cloud services and software container engines (18 schedulers/cloud services and 7 container engines, as of version 24.10.4) (Additional File 1: Supp. Table 6). These developments have been complemented by community-driven projects meant to expand Nextflow functionalities for the benefit of any pipeline. These include nf-test [23], a testing framework for Nextflow pipelines and components; and various Nextflow plugins such as nf-validation [24], a JSON Schema validation of pipeline parameters and sample sheets; nf-co2footprint [25], a monitoring tool for energy consumption and carbon emission from pipeline computations; and nf-prov [26], a provenance report tool for pipeline runs.

Impact of Nextflow module support on nf-core

By convention, a module is implemented as a Nextflow process definition around a specific functional task—typically an existing bioinformatics algorithm or script. An nf-core module is a Nextflow module developed according to nf-core standards and can be programmatically installed alongside the pipelines that use it. The standards involve declaring the required software dependencies via container images and Conda packages, including popular community-curated collections such as Bioconda [27] and Biocontainers [28]. In this way, each nf-core module encapsulates the correct environment for its embedded software, thus ensuring that each software works as intended. While this approach may stop short of solving incompatibilities caused by versioning, it nonetheless minimizes issues caused by two or more components being unable to coexist within the same environment.The transition to DSL2 has enabled the establishment of a shared repository for standardized modules and subworkflows. The community's strong adoption of this new modular approach is well illustrated by the rapid growth of the *nf-core/modules* repository, which has accumulated over 1400 modules and around 80 subworkflows in less than 5 years. As expected, a large number of modules were incorporated in the year the repository was created (> 700 in 2022), but the rate of new additions has remained significant, with over 300 and over 400 new entries in 2023 and 2024, respectively (Additional File 1: Supp. Table 6). Modules and subworkflows from the nf-core repositories can be used equally by both nf-core and non-nf-core pipelines. Beyond the public module repositories, nf-core also supports self-hosted module repositories. These allow for the direct integration of internally developed software with community-maintained modules and subworkflows, effectively bridging closed- and open-source development.

Best practices and improved pipeline standardization

The best practices are documented in publicly accessible online resources (Additional File 1: Supp. Table 4). They essentially revolve around ensuring that nf-core pipelines' implementations follow the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) [29, 30], good documentation practices, and precise conventions on code implementations. The best practices also rely on the incorporation of shared features applied across all nf-core pipelines, thus enabling consistent functionalities. To ease the use of its best practices, nf-core has developed a base template for pipelines. This template receives regular updates, which are semi-automatically synchronized across all pipelines. The template establishes a pipeline file structure with code, documentation, and continuous integration (CI) tests. Pipelines are tested using automated linting tests to check code formatting, syntax, and flag outdated nf-core modules. To guarantee deployability, each pipeline comes with minimal datasets used to test the pipeline for every pull request. CI testing also includes a "full-size" dataset with a realistic input to automatically test portability on AWS and Microsoft Azure before each stable release. The output files are publicly accessible and can be browsed on the nf-core website.

Extended tooling

The *nf-core/tools* package provides helper tools to improve accessibility for users of any Nextflow pipeline. Its functionality features commands to download pipelines and associated software for offline use, and methods for launching pipelines, including both a command-line and a graphical web-browser-based pipeline launch interface to support users less comfortable with command-line interfaces. The CLI provides features to create new pipelines based on the nf-core template and to produce and modify JSON schema files for parameter validation and documentation. Subcommands allow developers to discover, install, and update modules and subworkflows, as well as create new ones from a template.

Reproducibility and data provenance

Achieving full reproducibility when running a pipeline is an essential requirement of the FAIR paradigm. Nextflow and nf-core simplify this task by supporting the programmatic collection of execution parameters (e.g., tool versions, command-line flags) through parameter files, the nf-prov plugin, and MultiQC [31] summary reports. Collecting this data unambiguously defines the procedures used to generate the considered results, ensuring transparency and traceability. While the use of containerized pipelines enhances deployability and reproducibility, documenting all deployment parameters enables researchers to reproduce analyses or assess comparability between similar analyses. Provenance and the insights it provides on an analysis are especially critical when comparing results collected across large consortia, such as EuroFAANG (see the next section). This aspect is central within a consortium formally committed to meeting community-agreed metadata standards (like those of FAANG [32]) by facilitating comparability, reproducibility, and standardization.

Uptake of Nextflow and nf-core by research communities: the case study of the farmed animals genomics community

Over the last 7 years, Nextflow has been adopted by a diverse set of research communities (Additional File 1: Supp. Table 7). The nf-core community was the first one to emerge. It originated from the early Nextflow users and naturally structured itself around specific computational pipelines, bringing together scientists scattered across projects and institutions but connected by their common interest in a specific pipeline, whether as users or developers. Additional communities formed around shared scientific goals rather than a specific tool. Some of these research communities have developed outside the realm of nf-core. For instance, Genomics England (GE) was established by the UK's Department of Health with the long-term goal of providing the benefits of personalized genomics to the English population. This initiative began with the extensive sequencing of patients with rare diseases or cancer—already over 100,000 genomes from nearly 90,000 patients as of Data Release Version 19 [33]—and is now extending efforts towards sequencing 100,000 newborns within the Generation Study. GE has so far generated over 21 petabytes of data [34] and has become a central hub for the English medical community dedicated to patients whose conditions may be improved using genomic analysis. GE's key computational resource is a collection of Nextflow pipelines specifically dedicated to post-genomic analysis (Additional File 1: Supp. Table 7). Another community example is the Darwin Tree of Life (DToL) [35] hosted by the Wellcome Sanger Institute. DToL is the British initiative related to the Earth BioGenome Project (EBP) [36], whose objective is to sequence up to 2 million species around the world. DToL specifically aims to enable the sequencing of 70,000 species endemic to the British Isles and is one of the most advanced national chapters of the EBP. The DToL brings together a very interdisciplinary group of biologists sharing an interest in this massive sequencing effort. Just like for Genomics England, the DToL computational common denominator is a collection of high-quality pipelines dedicated to genomics analysis. These pipelines are not part of the nf-core collection, but they have all been explicitly developed using the nf-core standards and tooling (Additional File 1: Supp Table 7). Communities can also be united by a technology rather than a scientific question. This is the case of the Oxford Nanopore Technologies EPI2ME platform. EPI2ME supports scientists using nanopore sequencing technology [37]. EPI2ME provides this community with 15 curated open-source pipelines, all developed and documented to a standard comparable to nf-core (Additional File 1: Supp. Table 7). Their availability in the GitHub repository makes them amenable to any kind of customization.

These three examples illustrate how the use of common pipeline standards may act as a synergy catalyst. However, they also reflect contexts where well-funded, centralized organizations like Oxford Nanopore or Genomics England are able to provide dedicated support to their user communities. As inspiring as they may be, these examples are of little help for groups of scientists whose needs are not yet covered and who may not have adequate resources to collectively transition towards Nextflow or nf-core. Such groups need to weigh the costs and benefits and, most importantly, to evaluate the implications and feasibility of a slow transition, bearing in mind that partial transition and production may need to coexist for undetermined periods of time. For such an audience, the recent incorporation of the nf-core Animal Genomics Special Interest Group (Additional File 1: Supp. Table 7), one of the five new nf-core Special Interest Groups, constitutes a more telling example. We outline below the process through which a community featuring 80 research labs across Europe, the Middle East, Australia, and North America collectively organized a progressive standard transition towards the implementation and deployment of nf-core pipelines.

Functional Annotation of Animal Genomes project (FAANG)

FAANG [38] is a coordinated international effort established in 2014 with the purpose of harmonizing genotype-to-phenotype research in farmed animals. EuroFAANG is the European chapter of FAANG, developing as a future European Research Infrastructure [39] intended to connect existing pan-European infrastructures for data management, animal agriculture, and aquaculture. It also addresses the research priorities outlined in "FAANG to Fork" [40], which mirrors the European Commission's European Green Deal "Farm to Fork" strategy on aspects related to animal genetic resources, phenotyping and breeding, and animal health. The adoption, adaptation, and reuse of nf-core workflows for future genotype-to-phenotype research align with the goal of the

EuroFAANG Research Infrastructure to ensure that analyses conducted across European institutes and industry are interoperable, standardized, and reusable.

At the time this report was written, EuroFAANG encompassed six consortia of the EU's Horizon 2020 program: AQUA-FAANG, BovReg, GENE-SWitCH, RUMIGEN, GEroNIMO, and HoloRuminant. While each project specialized in a distinct set of farmed animal species (multiple fish species, cattle, pig, and chicken), they shared the common goal of improving our understanding of the genome-phenome relationship through the comprehensive annotation of genomic functional features (e.g., dynamic gene expression, chromatin accessibility, and epigenetic state under a range of biological conditions). Comprehensive functional annotation demands the generation and analysis of large sequencing datasets derived from distinct functional assays, including RNA-seq, ATAC-seq, and ChIP-seq. Together, these projects are committed to implementing an analysis strategy adhering to the FAIR principles of scientific reproducibility.

Necessity of interoperability and standardization

The use of nf-core provides an effective standardization strategy, which avoids technical differences potentially introduced by uncoordinated parameterization and variations across analysis tools. Such standardization ensures the comparability of results. For instance, AQUA-FAANG relied on nf-core pipelines to investigate the evolutionary conservation of genomic regulatory elements and epigenetic states across six farmed fish species for multiple biological conditions and matched tissue panels [41]. Since these conserved regions are known to be functionally important, their identification could be key to improving breeding strategies. Data integration, beyond the goal and lifetime of projects such as EuroFAANG, is also critical across time and constitutes an important part of sustainability. For instance, the FarmGTEx aims to characterize tissue transcriptomic profiles of farm animals, akin to its human counterpart project, GTEx [42–44]. As one would expect, this project will keep integrating new species. Yet, with sequencing and other techniques evolving, the interoperability of data collected over the years can only be guaranteed using a common best-practices analytic framework such as nf-core.

Main hurdles towards a common bioinformatics analysis framework

Despite the obvious rationale and good intentions across the board, establishing common standards remains a very challenging bioinformatics goal [45]. Research groups often rely on established methods, and adopting new standards requires retraining staff as well as refactoring perfectly functional pipelines. This complicated process can benefit from being spearheaded by a designated leader. For instance, when the AQUA-FAANG consortium decided to utilize nf-core, the University of Edinburgh and the Norwegian University of Life Sciences took responsibility for deploying the framework and training the other AQUA-FAANG groups to use nf-core pipelines. Adoption can also be achieved through recommendations, as demonstrated in GENE-SWitCH, whose initial proposal explicitly suggested Nextflow as a common WfMS for the entire consortium while leaving the implementation of this recommendation to each partner [46]. Regardless of which framework is chosen, the transition roadmap is as important as the endpoint. Resource-limited groups may benefit from a gradual transition to standardized pipelines, allowing new and legacy workflows to coexist during the adaptation period.

We argue in the next section that the continuum between low-level scripting, Nextflow, and nf-core has allowed EuroFAANG to go through a progressive standard transition that has allowed the coexistence of alternative data processing methods without compromising their long-term integration.

Choosing a bioinformatics analysis framework

Discussions and surveys about methods, languages, and standards were carried out among the EuroFAANG partners (Additional File 1: Supp. Table 8). These consultations indicated that Nextflow was the favorite WfMS and identified methodological gaps in nf-core. We found that the 15 consortium partners involved in data analysis were carrying out 78 analyses using four different programming languages: Nextflow (70), Snakemake (3), Bash (4), and Perl (1). Nextflow being the most widely used, its adoption provided the simplest paths toward harmonization. On top of this, Nextflow was designed to be language agnostic and, as a consequence, any existing scripts and functions of a pipeline can be directly ported into Nextflow in their original language (like Bash and Perl). The language-agnostic property may be achieved with other WfMSs through system calls. Nextflow implements idempotent task execution, a concept also used in other systems, which enhances reproducibility by uniquely identifying each task based on its inputs and code. This result is achieved by assigning each task a unique, repeatable identity based on a hash of its inputs and script and executing it in a dedicated work directory. Language agnosticism enables pipelines to be gradually ported into Nextflow without the need to rewrite scripts or do extensive revalidation. For example, the BovReg eQTL analysis workflow (Additional File 1: Supp. Table 8) was initially written in Bash and ported into a Nextflow pipeline with minimal refactoring and revalidation. The support for simple refactoring is essential in the case of limited resources.

EuroFAANG pipeline development within the nf-core ecosystem

All nf-core pipelines are open source and welcome new contributions. For instance, EuroFAANG partners have been actively involved in the maintenance and development of several existing nf-core pipelines, including *nf-core/atacseq*, *nf-core/chipseq*, and *nf-core/smrnaseq* (Additional File 1: Supp. Table 8). PacBio-based genome annotation was identified as a gap for the GENE-SWitCH project, leading one of its partners (Roslin Institute, University of Edinburgh, UK) to port their Iso-Seq Nextflow pipeline into nf-core (*nf-core/isoseq* [47]). Similar collaborative efforts have been initiated within GEroNIMO and BovReg for eQTL analysis using GWAS and Allele-Specific Expression (ASE). This led to the development of *nf-pegASE* (INRAE; Additional File 1: Supp. Table 8), a pipeline for ASE analysis that is based on the nf-core template, which is intended to be contributed to nf-core. Some Nextflow pipelines were also developed outside of the nf-core umbrella, like TAGADA [48] for the annotation and quantification of novel long non-coding and coding RNA transcripts and *BovReg/nf-cage* [49] for transcript annotation. Thanks to the standardization provided by Nextflow, these non-nf-core pipelines can nonetheless provide new modules and elements lending themselves to nf-core integration. Overall, most pipeline methods were widely shared across the various consortium members, and while nf-core dominates, other solutions were also part of a very active exchange process (Fig. 2B; Additional File 1: Supp. Table 8) with 4 of the 17 pipelines being used by at least three EuroFAANG consortia.

Conclusion and perspective

Nextflow and nf-core have become an essential component of the large-scale computational analysis on which modern genomics relies. A growing number of communities are now organized around this framework. The nf-core community is the largest Nextflow-centered community, and its rapid growth illustrates the attractiveness of strict best-practice guidelines bundled with tooling for developers and users. The new module and subworkflow infrastructure facilitates collaborative development both within nf-core and between external pipelines, leading to improved code quality and long-term maintenance. In the EuroFAANG context, and especially among BovReg partners, the use of nf-core and the availability of off-the-shelf pipelines ready to be adapted has allowed the development of additional pipeline resources within the initial budget. This unusual situation illustrates how the use of nf-core standards and community support enables faster development through improved interoperability and standardization. As adoption spreads from traditional bioinformatics to other scientific fields, these standards continue to enable other distributed communities to work together effectively and produce more reliable research.

Above we have described the nature of the internal exchanges that led a group of consortia to adopt a standard for bioinformatics pipeline development. Quantifying this adoption will require a precise bibliometric survey focused on nf-core-related practices, as reflected by citations. Given the inertia associated with publication, we anticipate this process to yield definitive results about 2 to 4 years from now. While this bibliometrics would provide the most robust estimator of nf-core stakeholder satisfaction, other data indicate their comfort with Nextflow, especially the *State of the Workflow* report (Additional File 1: Supp. Table 4) in which a survey conducted on 608 Nextflow users found 94% of them satisfied with Nextflow (2 points more than the 2023 survey). Although these surveys are informative, they only scratch the surface of Nextflow's demographic complexity. From their very start, the dynamic of Nextflow and nf-core has been a bottom-up process. None of these resources were officially commissioned by large grants, flagship projects or powerful institutions. They emerged at the local level and spread through word-of-mouth. Later, significant resources were committed to expanding the community in the most inclusive way, partly with the support of the Chan Zuckerberg Initiative [50, 51] and the institutes where Nextflow and nf-core were founded such as the Centre for Genomic Regulation (CRG, Spain), the Quantitative Biology Center (QBiC, Germany) and Science for Life Laboratory (SciLifeLab, Sweden). As data keeps accumulating, the time will soon be ripe to precisely quantify the effectiveness of this deployment process and explore the structure of the community, if only to understand what its future dynamics may be.

We also advocate here for progressive standard transitions at the level of research consortiums. We describe this process as taking place within a framework that accommodates a continuum between the current state of an implementation and the most desirable one, i.e., the standard. We found nf-core to allow a good trade-off between these two points, leveraging Nextflow's capability to encapsulate any existing code regardless of implementation language and deploy the execution at scale in a portable and replicable manner across heterogeneous execution platforms. Thus, strict adherence to nf-core best practices is not mandatory for a Nextflow pipeline to be functional, and the porting of any functional Bash pipeline into Nextflow can be achieved in a variety of ways, ranging from limited refactoring (i.e., wrapping of an existing pipeline with Nextflow tooling) up to full-scale refactoring. In our opinion, the extra development effort is offset by secondary benefits resulting from community input: nf-core pipelines are more visible and accessible, as evidenced by the statistics collected on the WorkflowHub, which already features 98 out of the 112 nf-core pipelines. The nf-core pipelines also benefit from community support, validation, bug fixes, and extension. These factors are especially important when considering long-term sustainability. All of these aspects are effectively demonstrated in the EuroFAANG collaboration. We expect the EuroFAANG developments to be merely a start and that the rapidly expanding Special Interest Groups initiative will help structure a growing number of communities around nf-core. We are optimistic that these groups will work orthogonally to those working on specific pipelines, promoting improved collaboration and standardization between scientists worldwide.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-025-03673-9.

Additional file 1: Supplementary_Material. Description: This file includes supplementary materials that support the findings in the main manuscript. Supplementary Figure 1 visualises [LE1] Scopus citation data (related to Supplementary Table 2) for bioinformatics workflow management systems like Galaxy, Nextflow, and Snakemake. Supplementary Tables 1 and 2 present citation counts from Google Scholar and Scopus, respectively, for the major publications of recently reviewed workflow management systems [6–8, 18, 52–71]. Additionally, this file contains other tables (Supplementary Tables 3 to 8) detailing metrics for workflow management systems, online resources, community outreach, development timelines, and examples of analyses in EuroFAANG projects. Furthermore, the Supplementary Section DSL2 is included. corr25_e5cccb68-2504-4441-b713-8633d8318f35

Additional file 2: Point-By-Point-Response-to-Reviewers.pdf. Description: This file contains the reviewer comments and the corresponding detailed responses.

Authors' contributions

Original draft and writing: B.E.L., J.E.-C., C.N., F.B., P.A.E., J.A.F.Y., E.W.F., G.Ga., M.U.G., C.H., F.H., M.H., E.M., J.M.-P., S.N., H.P., A.P., and M.d.C.R.-D. Re-writing and revision cycle: all authors. Bibliometric data compilation: B.E.L., J.E.-C., and C.N. Nextflow DSL2 description: P.D.T. Data collection (pipeline table): B.E.L., J.E.-C., A.A., M.-O.B., M.C., P.K.C., E.L.C., S.D., S.E., D.F., S.F., G.Gi., M.K.G., C.G., P.W.H., C.Ku., C.Kü., S.L., D.L., D.J.M., G.C.M.M., F.P., Y.R.-C., D.R., M.S., and A.S. nf-core metrics data: J.M.-P., P.A.E., C.H., F.H., M.H., E.M., M.d.C.R.-D., and J.E.-C. Sankey plot data: J.E.-C. Time graph data: nf-core community All authors read and approved the final manuscript.

Peer review information

Andrew Cosgrove was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Funding

The EuroFAANG genomics communities have received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement Numbers 817923 (AQUA-FAANG), 815668 (BovReg), 817998 (GENE-SWitCH), 101000226 (RUMIGEN), 101000236 (GEroNIMO), 101000213 (HoloRuminant), and the Horizon Europe programme under Grant Agreement Number 101094718 (EuroFAANG Research Infrastructure). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Horizon 2020 research and innovation programme,European Union,815668 (BovReg),815668 (BovReg),817923 (AQUA-FAANG),815668 (BovReg),815668 (BovReg),815668 (BovReg),817998 (GENE-SWitCH),101000236 (GEroNIMO),815668 (BovReg),817998 (GENE-SWitCH),817923 (AQUA-FAANG),817923 (AQUA-FAANG),817998 (GENE-SWitCH),815668 (BovReg),817998 (GENE-SWitCH),815668 (BovReg),817998 (GENE-SWitCH),817923 (AQUA-FAANG),817923 (AQUA-FAANG),815668 (BovReg),101000236 (GEroNIMO),101000213 (HoloRuminant),815668 (BovReg),815668 (BovReg),815668 (BovReg),815668 (BovReg),815668 (BovReg),Spanish Ministry of Science and Innovation,Generalitat de Catalunya,Werner Siemens-Stiftung,Palaeobiotechnology,Horizon Europe programme,European Union,101094718 (EuroFAANG Research

Infrastructure),101094718 (EuroFAANG Research Infrastructure),Carl-Zeiss-Stiftung,Certification and Foundations of Safe Machine Learning Systems in Healthcare,Deutsche Forschungsgemeinschaft,EXC 2180—390900677 (iFIT)

CN, BL and JEC have received funding from the Spanish Ministry of Science and Innovation. They acknowledge support from the Spanish Ministry of Science and Innovation through the Centro de Excelencia Severo Ochoa (CEX2020-001049-S, MCIN/AEI/https://doi.org/10.13039/501100011033), and the Generalitat de Catalunya through the CERCA programme. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement 815,668 (BovReg). They are grateful to the CRG Core Technologies Programme for their support and assistance in this work.

SN acknowledges support by the Carl Zeiss Foundation, project "Certification and Foundations of Safe Machine Learning Systems in Healthcare." SN is "gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder—EXC 2180—390,900,677 (iFIT)." SN is "gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder—EXC 2124–390,838,134 (CMFI)." This study was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) via the project NFDI 1/1 "GHGA—German Human Genome-Phenome Archive" (#441,914,366 to SN).

J.A.F.Y. was supported by Max Planck Society and the Werner Siemens Foundation grant "Palaeobiotechnology" (awarded to Profs. Pierre Stallforth and Christina Warinner).

Data availability

No datasets were generated or analyzed during the current study.

Declarations

Competing interests

Several Seqera employees are active promoters of nf-core. Seqera also provides direct infrastructure support (e.g. Wave) to the nf-core project as well as direct support to Nextflow and nf-core development. The Centre for Genomic Regulation and CN hold Seqera stocks and phantom stocks respectively. The CN lab at the Centre for Genomic Regulation receives support from Seqera.

Author details

[1]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain. [2]MED. University of Évora, Polo da Mitra, Évora 7002-554, Portugal. [3]Centre for Interdisciplinary Research in Animal Health (CIISA), Faculty of Veterinary Medicine, University of Lisbon, Lisbon 1300-477, Portugal. [4]Centre for Integrative Genetics (CIGENE), Faculty of Biosciences, Norwegian University of Life Sciences, Ås 1430, Norway. [5]Kungliga Techniska Högskolan, School of Engineering Sciences in Chemistry, Biotechnology and Health at the Department of Gene Technology, Stockholm, Sweden. [6]INRAE, AgroParisTech, GABI, Université Paris-Saclay, Jouy-en-Josas F-78350, France. [7]SIGENAE, INRAE, Jouy-en-Josas F-78350, France. [8]Research Institute for Farm Animal Biology (FBN), Wilhelm-Stahl-Allee 2, Dummerstorf 18196, Germany. [9]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Campus, Roslin, UK. [10]Seqera, Carrer de Marià Aguiló, 28, Barcelona 08005, Spain. [11]IRSD, Université de Toulouse, INSERM, INRAE, ENVT, Univ Toulouse III - Paul Sabatier (UPS), Toulouse, France. [12]GenPhySE, Université de Toulouse, INRAE, ENVT, Castanet-Tolosan 31326, France. [13]Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Pl. 6, Leipzig, Saxony 04103, Germany. [14]Current address: Department of Paleobiotechnology, Leibniz Institute for Natural Product Research and Infection Biology Hans Knöll Institute, Adolf-Reichwein-Straße 23, Jena, Thuringia 07745, Germany. [15]Current Address: Associated Research Group of Archaeogenetics, Leibniz Institute for Natural Product Research and Infection Biology Hans Knöll Institute, Adolf-Reichwein-Straße 23, 07745 Jena, Thuringia, Germany. [16]Natural Resources Institute Finland (Luke), Applied Statistical Methods, Myllytie 1, Jokioinen 31600, Finland. [17]Department of Pathology, Yale School of Medicine, New Haven, CT 06511, USA. [18]Biomedical Data Science, Department of Computer Science, University of Tübingen, Tübingen, Germany. [19]Quantitative Biology Center (QBiC), University of Tübingen, Tübingen, Germany. [20]M3 Research Center, Faculty of Medicine, University of Tübingen, Tübingen, Germany. [21]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK. [22]Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden. [23]SciLifeLab, Uppsala University, Uppsala, Sweden. [24]Friedrich-Loeffler-Institut (FLI), Greifswald-Riems 17493, Germany. [25]PEGASE, INRAE, Institut Agro, Saint Gilles 35590, France. [26]Department of Biological Sciences and Center for Systems Biology, The University of Texas at Dallas, Richardson, TX 75205, USA. [27]Unit of Animal Genomics, GIGA & Faculty of Veterinary Medicine, University of Liège, Liège, Belgium. [28]Institute for Bioinformatics and Medical Informatics (IBMI), University of Tübingen, Tübingen, Germany. [29]Boehringer Ingelheim Pharma GmbH & Co KG, Birkendorfer Straße 65, Biberach/Riss 88400, Germany. [30]Animal Breeding and Genetics Program, Institute of Agrifood Research and Technology (IRTA), Caldes de Montbui, Barcelona, Spain. [31]Postgraduate Program in Biotechnology at Universidade Potiguar (UnP), Natal/RN, Brazil. [32]Dairy Research Innovation Centre, South and West Faculty, Scotland's Rural College, Dumfries DG1 3NE, UK. [33]Universitat Pompeu Fabra (UPF), Barcelona, Spain.

Received: 30 May 2025 Accepted: 27 June 2025

References

1. THE Gtex CONSORTIUM. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369(6509):1318–30.
2. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74.

3. Kawaji H, Kasukawa T, Forrest A, Carninci P, Hayashizaki Y. The FANTOM5 collection, a data series underpinning mammalian transcriptome atlases in diverse cell types. Sci Data. 2017;4(1):170113.
4. Grüning B, Chilton J, Köster J, Dale R, Soranzo N, Van Den Beek M, et al. Practical computational reproducibility in the life sciences. Cell Syst. 2018;6(6):631–5.
5. Deelman E, Peterka T, Altintas I, Carothers CD, Van Dam KK, Moreland K, et al. The future of scientific workflows. Int J High Perform Comput Appl. 2018;32(1):159–75.
6. The Galaxy Community. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. Nucleic Acids Res. 2024;52(W1):W83-94.
7. Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, et al. Sustainable data analysis with Snakemake [version 2; peer review: 2 approved]. F1000Research. 2021 Apr 19;10(33). Available from: https://f1000research.com/articles/10-33. Cited 2024 Jan 24.
8. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. Nat Biotechnol. 2017;35(4):316–9.
9. Wratten L, Wilm A, Göke J. Reproducible, scalable, and shareable analysis pipelines with bioinformatics workflow managers. Nat Methods. 2021;18(10):1161–8.
10. About GitHub. GitHub. 2025. Available from: https://github.com/about. Cited 2025 Feb 27.
11. Gustafsson OJR, Wilkinson SR, Bacall F, Soiland-Reyes S, Leo S, Pireddu L, et al. WorkflowHub: a registry for computational workflows. Sci Data. 2025;12(1):837. https://doi.org/10.1038/s41597-025-04786-3.
12. Crosswell LC, Thornton JM. ELIXIR: a distributed infrastructure for European biological data. Trends Biotechnol. 2012;30(5):241–2.
13. Galaxy | Tool Shed. Available from: https://toolshed.g2.bx.psu.edu/. Cited 2025 Feb 27.
14. Snakemake workflow catalog. 2024. Available from: https://snakemake.github.io/snakemake-workflow-catalog/. Cited 2024 Jun 6.
15. Snakemake-Workflows. 2024. Available from: https://github.com/snakemake-workflows. Cited 2024 Feb 20.
16. nf-core. 2024. Available from: https://nf-co.re/. Cited 2024 Jun 6.
17. Grayson S, Marinov D, Katz DS, Milewicz R. Automatic Reproduction of Workflows in the Snakemake Workflow Catalog and nf-core Registries. In: Proceedings of the 2023 ACM Conference on Reproducibility and Replicability. Santa Cruz CA USA: ACM; 2023. p. 74–84. Available from: https://dl.acm.org/doi/10.1145/3589806.3600037. Cited 2023 Jul 5.
18. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol. 2020;38(3):276–8.
19. Scott DR, Cho H, Day CK, Deller AT, Glowacki M, Gourdji K, et al. CELEBI: the CRAFT effortless localisation and enhanced burst inspection pipeline. Astron Comput. 2023;1(44):100724.
20. Lehmann F, Frantz D, Becker S, Leser U, Hostert P. FORCE on Nextflow: Scalable Analysis of Earth Observation Data on Commodity Clusters. In: Cong G, Ramanath M, editors. Proceedings of the CIKM 2021 Workshops. Gold Coast, Queensland, Australia: CEUR-WS.org; 2021. (CEUR Workshop Proceedings; vol. 3052). Available from: http://ceur-ws.org/Vol-3052/short12.pdf.
21. Kanakaris N, Karacapilidis N. Predicting prices of Airbnb listings via graph neural networks and document embeddings: the case of the island of Santorini. Procedia Comput Sci. 2023;1(219):705–12.
22. Seqeralabs/wave. Seqera; 2024. Available from: https://github.com/seqeralabs/wave. Cited 2024 Feb 20.
23. Askimed/nf-test. Askimed; 2024. Available from: https://github.com/askimed/nf-test. Cited 2024 Feb 20.
24. Nextflow-io/nf-validation. nextflow-io; 2024. Available from: https://github.com/nextflow-io/nf-validation. Cited 2024 Feb 20.
25. Nextflow-io/nf-co2footprint. nextflow-io; 2024. Available from: https://github.com/nextflow-io/nf-co2footprint. Cited 2024 Feb 20.
26. Nextflow-io/nf-prov. nextflow-io; 2024. Available from: https://github.com/nextflow-io/nf-prov. Cited 2024 Feb 20.
27. Grüning B, Dale R, Sjödin A, Chapman BA, Rowe J, Tomkins-Tinch CH, et al. Bioconda: sustainable and comprehensive software distribution for the life sciences. Nat Methods. 2018;15(7):475–6.
28. da Veiga LF, Grüning BA, Alves Aflitos S, Röst HL, Uszkoreit J, Barsnes H, et al. BioContainers: an open-source and community-driven framework for software standardization. Bioinformatics. 2017;33(16):2580–2.
29. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data. 2016;3(1):160018.
30. Barker M, Chue Hong NP, Katz DS, Lamprecht AL, Martinez-Ortiz C, Psomopoulos F, et al. Introducing the FAIR principles for research software. Sci Data. 2022;9(1):622.
31. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047–8.
32. Harrison PW, Sokolov A, Nayak A, Fan J, Zerbino D, Cochrane G, et al. The FAANG Data Portal: Global, Open-Access, "FAIR", and richly validated genotype to phenotype data for high-quality functional annotation of animal genomes. Front Genet. 2021;12. Available from: https://doi.org/10.3389/fgene.2021.639238. Cited 2024 Jan 24.
33. Release v19 (31/10/2024) - Genomics England Research Environment User Guide. Available from: https://re-docs.genomicsengland.co.uk/release19/. Cited 2025 Feb 27.
34. Genomics England Case Study. WEKA. 2020. Available from: https://www.weka.io/resources/case-study/genomics-england-case-study/. Cited 2025 Feb 27.
35. The Darwin Tree of Life Project Consortium. Sequence locally, think globally: the Darwin Tree of Life Project. Proc Natl Acad Sci. 2022;119(4):e2115642118.
36. Lewin HA, Richards S, Lieberman Aiden E, Allende ML, Archibald JM, Bálint M, et al. The Earth BioGenome project 2020: starting the clock. Proc Natl Acad Sci. 2022;119(4):e2115635118.
37. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol. 2016;17(1):239.
38. Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW, et al. Coordinated international action to accelerate genome-to-phenome with FAANG, the functional annotation of animal genomes project. Genome Biol. 2015;16(1):57.

39. A European infrastructure for farmed animal genotype to phenotype research. 2022. Available from: https://cordis.europa.eu/project/id/101094718. Cited 2024 Jan 24.
40. Clark EL, Archibald AL, Daetwyler HD, Groenen MAM, Harrison PW, Houston RD, et al. From FAANG to fork: application of highly annotated genomes to improve farmed animal production. Genome Biol. 2020;21(1):285.
41. Johnston IA, Kent MP, Boudinot P, Looseley M, Bargelloni L, Faggion S, et al. Advancing fish breeding in aquaculture through genome functional annotation. Aquaculture. 2024;30(583):740589.
42. Liu S, Gao Y, Canela-Xandri O, Wang S, Yu Y, Cai W, et al. A multi-tissue atlas of regulatory variants in cattle. Nat Genet. 2022;54(9):1438–47.
43. Teng J, Gao Y, Yin H, Bai Z, Liu S, Zeng H, et al. A compendium of genetic regulatory effects across pig tissues. Nat Genet. 2024;56(1):112–23.
44. Guan D, Bai Z, Zhu X, Zhong C, Hou Y, Zhu D, et al. Genetic regulation of gene expression across multiple tissues in chickens. Nat Genet. 2025;57(5):1298–308. https://doi.org/10.1038/s41588-025-02155-9.
45. Leng R, Viry G, García-Sancho M, Lowe J, Wong M, Vermeulen N. The sequences and the sequencers: what can a mixed-methods approach reveal about the history of genomics? Hist Stud Nat Sci. 2022;52(3):277–319.
46. Watson M, Law A, Archibald A, Kuo R, Harrison P, Martin F, et al. Deliverable D2.1 Reproducible, scale-able workflows for FAANG data analysis; Available from: https://www.gene-switch.eu/uploads/1/2/6/6/126678735/deliverable_d2.1.pdf. Cited 2025 Feb 27.
47. Guizard S, Miedzinska K, Smith J, Smith J, Kuo RI, Davey M, et al. nf-core/isoseq: simple gene and isoform annotation with PacBio Iso-Seq long-read sequencing. Bioinformatics. 2023;39(5):btad150.
48. Kurylo C, Guyomar C, Foissac S, Djebali S. TAGADA: a scalable pipeline to improve genome annotations with RNA-seq data. NAR Genomics Bioinform. 2023;5(4):lqad089.
49. Salavati M, Espinosa-Carrasco J. MazdaX/nf-cage: nf-cage. Zenodo; 2022. Available from: https://zenodo.org/records/6855541. Cited 2024 Jan 25.
50. Nextflow and nf-core. Chan Zuckerberg Initiative. Available from: https://chanzuckerberg.com/eoss/proposals/nextflow-and-nf-core/. Cited 2025 Feb 10.
51. Nextflow and nf-core: Reproducible Workflows for the Scientific Community - CZI. Chan Zuckerberg Initiative. Available from: https://chanzuckerberg.com/eoss/proposals/nextflow-and-nf-core-reproducible-workflows-for-the-scientific-community/. Cited 2025 Feb 27.
52. The Galaxy Community. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. Nucleic Acids Res. 2022J 5;50(W1):W345–51.
53. The Galaxy Community. Correction to 'The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update'. Nucleic Acids Res. 2022;50(15):8999–8999.
54. Jalili V, Afgan E, Gu Q, Clements D, Blankenberg D, Goecks J, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. Nucleic Acids Res. 2020;48(W1):W395-402.
55. Jalili V, Afgan E, Gu Q, Clements D, Blankenberg D, Goecks J, et al. Corrigendum: The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. Nucleic Acids Res. 2020;48(14):8205–7.
56. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Čech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 2018;46(W1):W537–44.
57. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 2016;44(W1):W3-10.
58. Goecks J, Nekrutenko A, Taylor J. The Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010;11(8):R86.
59. Blankenberg D, Kuster GV, Coraor N, Ananda G, Lazarus R, Mangan M, et al. Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol. 2010;89(1):19.10.1-19.10.21.
60. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, et al. Galaxy: a platform for interactive large-scale genome analysis. Genome Res. 2005;15(10):1451–5.
61. Fillbrunn A, Dietz C, Pfeuffer J, Rahn R, Landrum GA, Berthold MR. KNIME for reproducible cross-domain analysis of life science data. J Biotechnol. 2017;10(261):149–56.
62. Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, et al. KNIME - the Konstanz information miner: version 2.0 and beyond. ACM SIGKDD Explor Newsl. 2009;11(1):26–31.
63. Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, et al. Data analysis, machine learning and applications. In 2008. p. 319–26.
64. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics. 2018;34(20):3600–3600.
65. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics. 2012;28(19):2520–2.
66. Sadedin SP, Pope B, Oshlack A. Bpipe: a tool for running and managing bioinformatics pipelines. Bioinformatics. 2012;28(11):1525–6.
67. Novella JA, Emami Khoonsari P, Herman S, Whitenack D, Capuccini M, Burman J, et al. Container-based bioinformatics with Pachyderm. Bioinformatics. 2019;35(5):839–46.
68. Lampa S, Dahlö M, Alvarsson J, Spjuth O. SciPipe: A workflow library for agile development of complex and dynamic bioinformatics pipelines. GigaScience. 2019;8(5):giz044.
69. Voss K, Auwera GV der, Gentry J. Full-stack genomics pipelining with GATK4 + WDL + Cromwell. F1000Research. 2017 8;6(1381). Available from: https://f1000research.com/slides/6-1381. Cited 2024 Jan 26.
70. Vivian J, Rao AA, Nothaft FA, Ketchum C, Armstrong J, Novak A, et al. Toil enables reproducible, open source, big biomedical data analyses. Nat Biotechnol. 2017;35(4):314–6.
71. Amstutz P, Crusoe MR, Tijanić N, Chapman B, Chilton J, Heuer M, et al. Common Workflow Language, v1.0. figshare; 2016. Available from: https://figshare.com/articles/dataset/Common_Workflow_Language_draft_3/3115156/2. Cited 2024 Jan 25.

Publisher's Note