technologies

*Article*

# Computational Resources and Infrastructures for a Novel Bioinformatics Laboratory: A Case Study

**Emanuel Maldonado** [1,2,*] **and Manuel C. Lemos** [1,2]

[1] RISE-Health, Department of Medical Sciences, Faculty of Health Sciences, University of Beira Interior, Av. Infante D. Henrique, 6200-506 Covilhã, Portugal; mclemos@fcsaude.ubi.pt
[2] CICS-UBI—Health Sciences Research Center, University of Beira Interior, 6201-506 Covilhã, Portugal
* Correspondence: e.maldonado@fcsaude.ubi.pt

## Abstract

Introduction: Bioinformatics is a relatively recent multidisciplinary research field continuously offering novel opportunities. Although many researchers are actively working in/with bioinformatics, some research centers still face difficulties in hiring bioinformaticians and establishing the appropriate (first) bioinformatics infrastructures and computational resources. In our research center, we started from scratch and established initial bioinformatics infrastructures for common use and also for the specific case of precision/personalized medicine. Case description: Here, we report a case study reflecting our specific needs and circumstances during the implementation of a novel bioinformatics laboratory. This involved the preparation of rooms, computer networks, computational resources novel designs, and upgrades to existing designs. Moreover, this work involved people from diverse areas and institutions, such as companies, institutional projects, informatics, and technical infrastructures services. Discussion and evaluation: The work resulted in the implementation of four novel designs dedicated to genomic medicine and in the adaptation of two existing designs dedicated to common use located in the dry-lab room. This is not an accurate and objective work, as it often depends on the available computer hardware and the target bioinformatics field(s). The four novel designs offered substantial improvements when compared to the upgraded designs, additionally corroborated by performance evaluations, which resulted in an overall highest performance of the novel designs. Conclusions: We present work that was developed over two years until completion with functioning infrastructure. This project enabled us to learn many novel aspects not only related to redundant disk technologies, but also related to computer networks, hardware, storage-management operating systems, file systems, performance evaluation, and also in the management of services. Moreover, additional equipment will be important to maintain and expand the potential and reliability of the bioinformatics laboratory. We hope that this work can be helpful for other researchers seeking to design their bioinformatics equipment or laboratories.

**Keywords:** bioinformatics infrastructures; bioinformatics laboratory; laboratory planning and management; bioinformatics computational resources; personalized medicine; genomic medicine; computing milieux; data management and integration; computer design and architecture; computer performance evaluation

Academic Editors: Manoj Gupta and Dennis Douroumis

Received: 13 March 2025
Revised: 6 June 2025
Accepted: 2 July 2025
Published: 4 July 2025

**Citation:** Maldonado, E.; Lemos, M.C. Computational Resources and Infrastructures for a Novel Bioinformatics Laboratory: A Case Study. *Technologies* **2025**, *13*, 285. https://doi.org/10.3390/technologies13070285

## 1. Introduction

Bioinformatics is a recent research field which was pioneered in the 1960s by Margaret Dayhoff [1] and had its nomenclature first assigned in 1971 [2]. Currently, many researchers are actively working in/with bioinformatics, yet research centers still face difficulties in hiring bioinformatics specialists (e.g., [3]). This poses additional difficulties when establishing the appropriate (first) bioinformatics infrastructures and computational resources. Moreover, the difficulties increase when bioinformatics comprises different areas of research and when having to select adequate computational resources, ideally by employing computer science skills (e.g., [3,4]). Our research center, the Health Sciences Research Center (CICS-UBI) [5], was lacking computational resources with sufficient capacity for the different bioinformatics needs and biological levels [5]. Here, we start from scratch in establishing initial bioinformatics infrastructures and computational resources for common use [6], and also for the specific case of precision/personalized medicine [3,7–13].

Bioinformatics often consists of software research and development, thus enabling the generation and analysis of biological omics data. Such processing often starts at the sequencing and assembly stages [10,11,14,15] and may include additional downstream analysis (e.g., [16–18]). Each analysis is dependent on the biological level being addressed, such as the genome, transcriptome, epigenome, metagenome, and even metabolome and proteome [10,12,14], also potentially integrated as part of multi-omics approaches [10,13,19–22]. Each case can, thus, be framed into the big data [8,11–13,15,18,23,24] 5Vs criteria—velocity (generation), volume (size), veracity (uncertainty), variety (multi-omics, health records, and imaging, among others), and value (for the health and biological sectors/sciences)—and requires knowledge for designing appropriate computational resources [18] such as, for instance, considering the Graphical Processing Unit (GPU) [25] cards required for molecular dynamics studies [23,26]. Such data analysis and processing typically require computational resources capable of high-throughput [6,16,27–30] (HTC) and high-performance [15,19,23,31–36] (HPC) computing. This project will, thus, focus on genomics needs, which are compatible, among others, with evolutionary genomics [16,29,30,37,38], genomic medicine [7,9,10,17,39–43], and microbial genomics [44–46]. Hereafter, we describe the starting conditions.

For the establishment of the bioinformatics laboratory (a project which started in March 2022), we had at our disposal (i) four initial spaces at CICS-UBI, (ii) five desktop computers acquired in 2019 (Table 1), and (iii) four furniture working stations. Both (ii) and (iii) were adapted from the terminating project Cloud Computing Competence Center (C4-UBI) [47] in August 2022. Moreover, this project was started while being involved in a genomic medicine funded project [7], which had as a main objective to also establish the computational resources to facilitate research in this field pertinent to our research team. Herein, additional novel computational equipment was designed and prepared over a year while in contact with several entities. This resulted in four desktops, one workstation, and two Network-Attached Storage (NAS) servers as part of a scalable data center.

**Table 1.** Informatics resource designs initially available at C4-UBI.

| Brand/Model | HP EliteDesk 800 G4 TWR | G.Skill |
|---|---|---|
| #Specimens | 4 | 1 |
| Year | 2019 | 2019 |
| MB | Intel Q370 PCH-H—vPro | ASUS TUF B450-PLUS GAMING |
| CPU | Intel Core i7-8700 (6C12T) | AMD Ryzen 7 2700X (8C16T) |
| RAM | Samsung 16 GB DDR4-2666 MHz | G.Skill AEGIS 64GB DDR4-2400MHz |
| Storage 1 | SSD NVMe Samsung PM981 256 GB | SSD SATA BlueRay M8S 240GB |
| Storage 2 | na | HDD Toshiba P300 2 TB 7200 rpm |
| GPU | NVIDIA GeForce GT 730 | NVIDIA GeForce RTX 2060 SUPER |
| ECC | ns | ns |

na—not available. ns—not supported. HP: Hewlett-Packard.

Considering the aforementioned resources, the space available at CICS-UBI (Supplementary Materials S1: Figure S1) consisted of (i) a common room for performing any bioinformatics related studies and research (dry-lab) containing a single desk, (ii) two offices for researchers and professors, (iii) a server room for the establishment of the NAS servers, and (iv) a wet-lab where the sample preparation techniques take place, including the validation of bioinformatics results.

Here, we report a case study focused on the implementation of infrastructures and design of computational resources in the context of bioinformatics data management and genomics data analyses. Other published works that report on bioinformatics infrastructures and computational resources have been found (e.g., [6,13,18,23,31,48–52]). However, to our knowledge, this is the first work of its kind to develop on the technical and computational details of preparing these resources adapted for bioinformatics needs and, thus, to attempt to optimize the working conditions from the start. This case study may be useful for anyone interested and, among others, as an example for bioinformatics, for computer science researchers or groups willing to implement similar resources, considering not only the computer architecture, but also the integration with facilities and similar companies locally available.

The remainder of this article is organized as follows: the Section 2 introduces the reasoning behind the process of developing the room adaptation and upgraded and novel designs, accompanied by scalability considerations, costs per design, and final costs at the end. The Section 3 presents a few general remarks, potential bioinformatics applications, performance evaluation, and a comparison with EMBL-EBI criteria for core facilities. The Section 4 summarizes what was accomplished and points to future development directions.

## 2. Case Description

The case involves the design and adaptation of resources taking into consideration the aforementioned existing conditions (rooms, furniture, and computers), with the final objective to achieve optimal working conditions for the bioinformatics laboratory and researchers (Supplementary Materials S1). This also included the scalability, enabling adaptation to novel data and processing requirements.

Throughout the work presented in this section, we have attempted to seek an equilibrium between maximizing the performance of the designs while, at the same time, maintaining low costs. To accomplish the implementation of our bioinformatics laboratory, a total of four acquisitions were necessary (please see the subsection Total and Final Costs). Hereafter, the costs for each computer design and adaptation are presented with value-added tax (VAT, 23%). All the costs presented in this section also include other components costs not listed in the following tables (Tables 2–8), such as Central Processing Unit (CPU) [25] dissipation fans, cables, chassis, adapters, and other materials, which can be found in more detail in the Supplementary Information (Supplementary Materials S2–S6). Beyond this, the Supplementary Information regarding Supplementary Materials S2–S6, concerns different countrywide companies who are mentioned hereafter not for publicity reasons, but in order for the reader to understand some of their differences. Effectively, they had an impact on the selected purchased material and were important to meet the needs of the different designs [53,54]. For instance, the company that provided the novel designs was not able to provide all the upgrades to the existing designs, as these would depend on the brand specific material, thus requiring us to seek an alternative. Additionally, based on these example companies, the interested reader is also able to identify similar services in their region for their own work [53,54].

This section is organized into the following three subsections: (i) Existing Resources and Adaptations (with two subsections), (ii) Novel Computational Resources (with five subsections), and (iii) Total and Final Costs. The first describes the room adaptation and upgrades to existing resources. The second describes the four novel designs with approximate costs, including additional material also important for (i). The third resumes the acquisitions by presenting the exact individual costs per acquisition and total expenditure.

### *2.1. Existing Resources and Adaptations*

Here, we will describe the existing conditions and adaptations at CICS-UBI, not only regarding the dry-lab room, but also considering furniture and the existing computational designs (Table 1).

#### 2.1.1. Dry-Lab Room Adaptation

The dry-lab had four network ports and six power outlets, which was clearly not enough for a room intended to host five desktops (Table 1) and a workstation (Table 7; please see subsection Novel Computational Resources). Hence, after a few measurements, it was found that this room (with an area of 19.20 m$^2$; Supplementary Materials S1: Figure S1) could host at least six desks. As the room had a single desk, the additional four from the C4-UBI project would enable us to prepare the room for five working stations and still leave room for an extra station that could be added later.

With this in mind, a plan was prepared and shared with the UBI technical infrastructure services (UBI-TIS), illustrating the conditions to be implemented in the room (Supplementary Materials S1: Figure S2). Specifically, for each working station, this plan initially foresaw the implementation of seven power outlets and of 14 network ports (Rejistered Jack-45—RJ45, each providing 1 Gigabit Ethernet—GBE speeds) distributed in pairs (one additional pair for a wall in case of the need for additional resources). After a visit from the UBI-TIS, it was found that this initial plan could be improved. Among other suggestions, it was proposed that Uninterruptible Power Supply (UPS) units could also be used, enabling us to reduce clutter and optimize space. The required material for fulfilling this plan was defined and prepared by the UBI-TIS and supplied by the company L3W Material Eléctrico Lda, Famalicão, Portugal (Supplementary Materials S5). At this moment, three UPS units (Riello NPW 1500VA/900W), each one with capacity for supporting two desktops (see also Table 1), were being prepared in parallel as part of the acquisition presented in the following subsection—Novel Computational Resources. Here, we considered the possibility of acquiring a single UPS unit for all desktops. However, knowing that UPS units will surely require repair or replacement, this solution would avoid affecting all the desktops. Beyond the three UPS units (EUR 616.23; Supplementary Materials S2: Proposals S1 and S2, page 5), the final room adaptation essentially consisted of (i) RJ45 network cables provided by the UBI coordinating informatics services (UBI-CIS) (without further costs), (ii) a fiber-optics cable with a length of 70 m and fiber-optics adapters (Table 8; Supplementary Materials S3 and S4), enabling the room connectivity to UBI infrastructures (EUR 124.29 + EUR 15.21), and (iii) among other technical materials, a small 19″ rack for housing the managed network switch (EUR 3508.14; Supplementary Materials S5). The fiber-optics cable was defined and prepared by the local technical informatics services (CICS-LTIS) and supplied by the company Copper2Fiber—Soluções de Conectividade, Lda, Barcarena, Portugal (Supplementary Materials S3). The fiber-optics adapters revealed necessary for enabling the connection of the fiber-optics cable to an intermediate switch (Supplementary Materials S1: Figure S2). These were defined by CICS-LTIS and supplied together with the computer upgrades (Tables 2 and 8; Supplementary Materials S4). Overall, the room was finalized, providing 14 network ports and 40 power outlets, half of which were adapted to provide electrical current from the three UPS units. The total cost of acquisition for all necessary material was EUR 3508.14 + EUR 616.23 + EUR 124.29 + EUR 15.21 = EUR 4263.87.

As the UBI-TIS performed the material installation and UBI-CIS/CICS-LTIS conducted the network equipment integration and activation, there are no further costs to be reported.

### 2.1.2. Existing Computational Designs Adaptation

Moreover, it was anticipated that the designs from Table 1 would require hardware upgrades, especially targeting Random Access Memory (RAM) and storage space. For this purpose, we consulted two companies, the one who originally supplied the four HP desktops and a second one (RuiPolana in collaboration with Eurobit—Sistemas Informaticos e Manutenção, Lda, both from Covilhã, Portugal) in order to find a competitive purchase and alternative options for all five desktops [53,54]. While consulting the first company, we attempted to understand the possibilities of additionally upgrading the CPU (replacing the current i7 for the i9-9900 from the same generation) and even including GPU options. In their reply, they supplied documentation from which we could understand that these options were not recommended, and, thus, we proceeded to upgrade RAM and storage.

From Table 2, it is possible to observe that all desktops will improve their specifications (Table 3—in italics). The total cost for this equipment was EUR 3453.42 (Supplementary Materials S4).

**Table 2.** Hardware upgrades for designs shown in Table 1.

| Desktop | Component | Specifications | Units (Number) |
|---|---|---|---|
| HP EliteDesk 800 G4 TWR | RAM | Kingston 16 GB DDR4-2666 MHz | 16 |
| | SSD NVMe | SSD NVMe WD M.2 PCIe 4.0 4 TB Black SN850X; | 4 |
| | | SSD WD M.2 PCIe 4.0 NVMe 1 TB Black SN850X | 2 |
| G.Skill | SSD NVMe | SSD NVMe WD M.2 PCIe 4.0 4 TB Black SN850X | 1 |
| | SSD SATA | SSD 2.5″ SATA SAMSUNG 1 TB 870 EVO | 2 |
| | Kit Keyboard + mouse | Logitech MK120 | 1 |

WD: Western Digital. HP: Hewlett-Packard.

**Table 3.** Resulting hardware upgrades in place for desktops shown in Table 1.

| Brand/Model | HP EliteDesk 800 G4 TWR | G.Skill |
|---|---|---|
| #Specimens | 4 | 1 |
| Year | 2019 | 2019 |
| MB | Intel Q370 PCH-H—vPro | ASUS TUF B450-PLUS GAMING |
| CPU | Intel Core i7-8700 (6C12T) | AMD Ryzen 7 2700X (8C16T) |
| RAM | *Kingston 64 GB DDR4-2666 MHz* | G.Skill AEGIS 64 GB DDR4-2400 MHz |
| Storage 1 | SSD NVMe Samsung PM981 256 GB | *2x SSD SATA Samsung 1 TB 870 EVO* |
| Storage 2 | *SSD NVMe WD 4 TB SN850X* | *SSD NVMe WD 4 TB SN850X +* HDD Toshiba P300 2 TB 7200 rpm |
| GPU | NVIDIA GeForce GT 730 | NVIDIA GeForce RTX 2060 SUPER |
| ECC | ns | ns |

ns—not supported. WD: Western Digital. HP: Hewlett-Packard.

It is clear from Table 3 that two SN850X 1 TB Solid-State Drives (SSDs) from Table 2 are not shown. These are being saved as backups in case some of the older 256 GB SSD drives from the HP units fail, thus have the possibility to immediately replace them. With this upgrade, the maximum RAM capacity was achieved for the HP desktops. Regarding the G.Skill desktop, the new drives were installed, offering a total capacity of 8TB, of which 6 TB is usable. The SSD from the G.Skill desktop was replaced in order to increase storage capacity and the original Hard Disk Drive (HDD) was maintained for temporary data. Considering the desktop age and specifications, an improvement in RAM capacity and frequency would not bring about much difference taking into account the potential added costs. The CPU RAM capacity limit is 64 GB and the RAM frequency is 2.9 GHz. However, when considering the limiting motherboard (MB) RAM frequency, this value drops to 2.6 GHz, thus without generating any improvements.

### *2.2. Novel Computational Resources*

While in the previous subsection, we described room planning and desktops upgrades, which took place during mid-semester of 2023, here, we present four designs for novel computational resources. This process started previously in March 2022 and while in a budget-free condition [7]. This was not possible without the specialized constant contact with a company (Nexus Solutions S.A., at the time named Soon—Business Solutions, S.A., Maia, Portugal). Nexus focuses on assembling informatics equipment nearly piece-by-piece based on a client's specific requirements, thus allowing for alternative solutions on the market. Additionally, the close contact with CICS-LTIS and UBI-CIS was important for providing important insights into the definition of these designs and appropriate integration into the aforementioned bioinformatics spaces, including the university local server room and network configurations, thus ensuring cybersecurity [55].

To provide a closer idea of the costs involved for each final design, they are presented as approximate to the acquisition (Supplementary Materials S2) and were obtained from the finalized designs of individual proposals from October 2022 (Supplementary Materials S6).

#### 2.2.1. NAS Servers—The Centralized Data Storage Solution

Here, we describe the design process and requirements for reaching the final genomic medicine data server designs.

Due to their complexity, the design and preparation of computational resources were started with the server architectures. As these were focused on data storage and access, this would also influence the storage available to the following designs, thus leading to designs with the minimum requirements. Here, two servers were designed, a main server (Data Center—DC) and a backup server (Data Center Backup—DCB).

The DC was foreseen to be continuously active, thus enabling the storage and access of two data types. One data type corresponded to the Next-Generation Sequencing (NGS) or genomics data, currently consisting of Whole-Exome Sequencing (WES) data generated from human samples, and the second data type, considered as "generic", corresponded to the Private Accounting Storage (PAS) for each researcher's personal work storage/backup, as the researcher responsible for their disk quota will be able to store any information he/she deems important from their research. Hence, among others, genomics, metadata, in-house software/scripts, and text/CSV files are all considered to be valid types. For each case and based on our requirements, we estimated a starting usable capacity of 40 TB for the NGS volume and no less than 18 TB for the PAS volume. In the former case, 40 TB would enable the storage of an estimated amount of 1500 WES data samples, with each sample being approximately 25–26 GB in size. In the latter case, this capacity could provide 1 TB of storage space for each element of the team (currently comprising seven elements) and possibly leave room for disk quota extensions and additional elements entering the team.

The DCB was foreseen to be online two times per month and was devised to provide backups of the DC server regarding the PAS data. A backup for the NGS data was decided to be unnecessary, since this data mostly arrives in external HDDs and/or SSDs. Even in the case where NGS data arrives via wire transfer (e.g., Amazon Web Services—AWS), additional external drives can be acquired for its storage. This solution enabled NGS data to be permanently safeguarded anywhere and recovered whenever needed. Moreover, this avoided an increase in costs regarding the backup solution, but placed more relevance on the PAS volume. Furthermore, the DCB server revealed to be necessary in order to avoid eventual failures, such as those regarding the Operating System (OS), any human errors, and any server components (e.g., HDDs) [56]. However, in order to perform the backups of the PAS data, both servers had to adopt a similar design (i.e., enclosure, number of HDDs, OS, Redundant Array of Independent Drives (RAID) scheme [57], and, thus, usable capacity). In this sense, both these designs were also prepared taking into consideration the use of UPS, which are mostly important for maintaining data integrity [56].

Both servers' designs were based on initial research and the selection of a more specialized operating system that could provide resilience and the more recent RAID technology. Here, we opted for the TrueNAS SCALE [58], a recent Debian-based system providing RAIDZ [56,59,60] and the newer implementations of declustered RAID (dRAID) [61–64] from its base Open Zettabyte File System (OpenZFS) [65,66]. Thus, offering large advantages in data and drive integrity [56,60–63]. The DC, through the TrueNAS SCALE, has the potential to be adapted to provide additional or alternative functionalities, including storage volumes for different types of omics data sets. At the lower level, the OpenZFS (since v.2.2) enables the following three alternative ways for pool or total capacity expansion: (1) by replacing all disks with higher capacity disks, (2) by increasing the number of vdevs/pools, or (3) more recently, it will be possible to expand a specific vdev/pool in any number of disks, including the foreseen case of dRAID levels [59]. Moreover, at a higher level, it can allow for the expansion of the server/node network [58]. This enables, in future, to acquire additional updated servers or workstations, which, while being interconnected, may increase the capacity for storage and data analyses [58].

To achieve these advantages, these OS and RAID schemes had some special requirements that needed to be satisfied [67]. This was especially true when considering the dRAID with double parity (dRAID2) scheme for both drive pools, NGS and PAS. Thus, the design process started with requesting an initial specification from the Nexus company, which resulted in considering the traditional hardware-based RAID5. However, this proposal was not accepted, as both the RAID5 and RAID6 were previously considered as obsolete [68,69]. After a few rounds of improvement, the final result was achieved. However, in an attempt to corroborate our designs, we sought feedback from users most experienced with TrueNAS systems. To this end, we submitted these designs to the consideration of the TrueNAS community forum [70]. Based on the discussion generated here, we further improved the designs by introducing the following four main changes: (i) doubling the RAM capacity (to 128 GB); (ii) replacing the MB with one providing a double CPU socket; (iii) replacing the single enclosure with 16slots with two separate enclosures with different sizes (namely, one with eight HDD slots for PAS data and one with 24 HDD slots for NGS data); and (iv) additionally, we were advised to use Serial Attached SCSI (SAS) expanders [67]. This kind of connectivity enables connections within longer distances between the Just-a-Bunch-of-Disks (JBOD) enclosure and the main disk controller, thus also enabling scalability. Changes to the DCB were only required to ensure that the CPU had more capacity to handle the associated RAID software processes and OpenZFS tasks. Hence, the CPU was modified from one with 4C4T (i.e., [n]Cores[m]Threads) to one with 6C12T. Both servers' designs are listed in Tables 4 and 5, respectively.

**Table 4.** Nexus DC server final design.

| Component | Specification | Units (Number) |
|---|---|---|
| Motherboard | Supermicro X12DPL-I6 (dual socket) | 1 |
| CPU | Intel Xeon Gold 6326 16C32T 2.9 GHz 24 MB | 1 |
| RAM | DIMM Samsung 32 GB DDR4 3200 MHz ECCR 2Rx4 (128 GB) | 4 |
| Boot drives (mirror) | SSD Samsung 240 GB PM893<br>SSD Samsung 240 GB SM883 | 1 + 1 |
| HBA controller | Broadcom 9500-8e Tri-Mode PCI-e 4.0<br>Broadcom 9500-8i Tri-Mode PCI-e 4.0 | 1 + 1 |
| Enclosure (main 8 HDD slots) | HDD Seagate Exos 7e8 6TB SATA3<br>HDD Seagate Exos 7e10 6TB SATA3 | 4 + 4 |
| Enclosure (JBOD 24 HDD slots) | HDD Seagate Exos 7e8 8TB SATA3<br>HDD Seagate Exos 7e10 8TB SATA3 | 5 + 5 |
| NIC | Broadcom P210TP dual port 10GBE BASE-T | 1 |
| UPS | Riello SDH 3000VA/2700W | 1 |
| *Extra* Boot drives | SSD Samsung 240 GB PM893<br>SSD Samsung 240 GB SM883 | 1 + 1 |
| *Extra* Storage drives (main 2 slots used) | HDD Seagate Exos 7e8 6 TB SATA3<br>HDD Seagate Exos 7e10 6 TB SATA3 | (*1* + *1*) * |
| *Extra* Storage drives (JBOD 2 slots used) | HDD Seagate Exos 7e8 8 TB SATA3<br>HDD Seagate Exos 7e10 8 TB SATA3 | (*1* + *1*) * |

(*) drives required and used for the final design. Units shown with plus sign signal the number of units per disk model.

**Table 5.** Nexus DCB server final design.

| Component | Specification | Units (Number) |
|---|---|---|
| Motherboard | Supermicro X12STL-F | 1 |
| CPU | Intel Xeon E-2336 6C12T 2.9 GHz 12 MB | 1 |
| RAM | Samsung 32 GB DDR4 3200 MHz ECC UNBUFFERED | 1 |
| Boot drives | Samsung 256 GB PM9A1 PCI-E 4.0 | 1 |
| HBA controller | Broadcom 9341-8i SATA/SAS3 | 1 |
| Enclosure (main 8 HDD slots) | HDD Seagate Exos 7e8 6 TB SATA3<br>HDD Seagate Exos 7e10 6 TB SATA3 | 4 + 4 |
| UPS | Riello SDH 2200VA/1980W | 1 |
| *Extra* Storage drives (main 2 slots used) | HDD Seagate Exos 7e8 6TB SATA3<br>HDD Seagate Exos 7e10 6TB SATA3 | (*1* + *1*) * |

(*) drives required and used for the final design. Units shown with plus sign signal the number of units per disk model.

The cost for each server was estimated to be EUR 16,249.78 for the DC (Supplementary Materials S6: Proposal S1) and EUR 5501.79 for the DCB (Supplementary Materials S6: Proposal S2). The extra eight drives shown in both Tables 4 and 5 had an additional total cost of EUR 1943.79 (Supplementary Materials S6: Proposal S3).

As seen from Table 4, the DC is composed of two enclosures. The JBOD is dedicated to the NGS volume and connected (via SAS expanders) to the smaller main enclosure dedicated to the PAS volume, additionally holding the server control and processing. The DCB (Table 5) is composed of a single enclosure similar to the DC main one.

Considering the above designs, the DC essentially offers four scalability possibilities, namely regarding the (i) MB: one more CPU can be installed, thus increasing the processing capacity, which could become useful in case the number of JBODs, storage pools, or drives increases; (ii) RAM: with the installation of an additional CPU, it becomes possible to increase the RAM capacity, and, at any time, the modules could be swapped for higher RAM capacity modules (limited to a total of 2 TB); (iii) HDD: the JBOD itself still provides room for additional 14x HDDs, enabling the creation of novel NGS volume(s) or expansion of the existing NGS volume capacity (see (1)–(3) above); and (iv) JBOD: due to the existence of SAS expander connectivity, it becomes possible to increase the number of JBOD enclosures.

Regarding the DCB, the scalability possibilities are essentially related to the RAM and PAS volume capacities. The maximum RAM capacity is limited to 128 GB, thus enabling the addition of up to three memory modules. Moreover, both PAS volumes could also increase their capacity, for instance, by replacing all HDDs with ones providing a higher capacity (see (1) above).

The designs shown in Tables 4 and 5 exhibit additional rows described as "Extra Boot/Storage drives". They were planned as contingency measures to fulfill any of the three following unforeseen cases (whichever occurred first): (i) a change in OS requirements, (ii) immediate replacement in the case of malfunctions or breakdowns, or even with (iii) a sudden increase in available data volume, thus requiring an increase in storage capacity and scalability [59]. Hence, Tables 4 and 5 show these rows with the quantities of parenthesized units (1 + 1). This number format pertains to each unit model and intends to convey that they later became part of the final designs (Tables 4 and 5). Here, the table rows named "Enclosure" consider both the initially selected drives plus the extra drives, thus showing the total units in use, 4 + 4 and 5 + 5, for each data volume. These changes resulted from tackling unforeseen needs regarding (i) above (please see the subsection Server RAID Configurations and Novel OS Requirements). They were necessary to ensure that the previously decided volume storage capacities were concomitant with the attainment of the most up-to-date RAID requirements from end-2023 releases of TrueNAS SCALE. This adjustment consisted of installing each pair of extra HDDs in the available two HDD slots of each enclosure. However, the main DC enclosure required additional attention, as the two HDD slots were not available as expected. Here, we found that these slots were being occupied by the two DC boot drives. Hence, the Nexus company was required to intervene (after acquisition and delivery) to provide room for the two extra 6 TB HDDs (Table 4). The solution consisted of moving them to a top layer of the enclosure. This caused the loss of the plug-and-play ability, which was expected to be possible in both servers from the start. This could have been avoided if, during design, we understood that the selected enclosure was not offering the additional slots for the two SSD boot drives, which we found to be possible on the back side of other similar 8x slots enclosures only during this difficulty resolution.

### 2.2.2. Servers RAID Configurations and Novel OS Requirements

The design of the DC and DCB servers specifically focused on attempting to harmonize all the remaining hardware components, with the intended OS TrueNAS SCALE and the underlying OpenZFS RAIDZ/dRAID levels [61,63] (hereafter generally referred to as ZFS RAID).

Initially, we planned to place all the pools working under a dRAID2 level [71]. This involved understanding that the HDDs would function with a higher performance by offering the 4 KB native (4Kn) advanced format (AF) [72–74]. This format, beyond being more adequate for larger file sizes, it could also offer advantages in terms of its inner records. Here, the Error Correction Code (ECC) fields provided an increased capacity [72,73], which can be important for improved ZFS RAID data integrity tasks [61,65,67,71]. Support for the ECC additionally required that the CPU, RAM, and MB were compatible with ECC specifications [67]. Moreover, due to ZFS RAID being a software-based RAID alternative, the definition of hardware components, such as the Host Bus Adapter (HBA) controller, RAM, and CPU capacity, was relevant for ensuring good functioning conditions. These components were also correlated with the number of HDDs in the system [67].

Regarding the configuration of dRAID2, the information initially available was not very accurate or abundant, but enabled to establish an approximate number of disks required. Hence, beyond the extra drives, we planned that the six HDDs for the PAS volumes and eight HDDs for the NGS volume would be sufficient. Hence, to achieve 18 TB, we established that six 6 TB HDDs could be distributed as one distributed hot-spare, two parity, and three data. Similarly, to achieve 40 TB, we established that eight 8 TB HDDs could be distributed as one distributed hot-spare, two parity, and five data disks. However, upon preparing to install the OS TrueNAS SCALE 23.10 providing the first support for dRAID [71], we encountered different requirements. Here, the number of data disks was stipulated differently than expected with a simple rule of power of two (2, 4, 8, and so on). Hence, by using the same number of HDDs, it became unfeasible to implement the dRAID2, especially in the PAS volumes. This then demanded to install the extra HDDs (Tables 4 and 5), thus achieving the maximum eight for the PAS volumes and ten for the NGS volume. Even though, with eight disks we had to resort to implement a more traditional RAIDZ2 level [65] in the PAS volumes. In the NGS volume, we could maintain the dRAID, but with single parity (dRAID1) [71]. Thus, this resulted in two RAIDZ2 levels with five data disks, two parity disks, and a single hot-spare and the dRAID1 with a single distributed hot-spare, single parity, and eight data disks [71]. These new configurations provided, respectively, 25 TB and 58 TB of usable capacities for each of the two PAS and NGS volumes. Although we may have lost some resilience by downgrading from dRAID2, on the other hand, additional storage capacity was gained free of cost.

### 2.2.3. Desktop Computers

In the case of the desktops, we opted to design a model desktop computer, ideally providing more adequate features enabling high performance and data integrity.

With this in mind, we selected the 12th generation CPU Intel i9-12900K, which was, at the time, the first generation to provide ECC capabilities ideal for ensuring data integrity. As these CPUs also provide the newest Double Data Rate (DDR) 5 memory standard, we planned DDR5 ECC RAM (64 GB) with a maximum frequency of 4.8 GHz. However, from the two MBs available (Supermicro X13SAE and X13SAE-F), despite being compatible with this memory type, their RAM clock frequency was limited to 4.4 GHz, hence limiting the design to this frequency. In terms of storage, we chose two SSD drives, where one with 500 GB would be sufficient for use as Linux/UNIX swap memory and the other SSD with 4 TB would be entirely used for the OS and data analyses. This design additionally considered monitors and UPS (Table 6).

**Table 6.** Nexus desktop final design.

| Component | Specification | Units (Number) |
|---|---|---|
| Motherboard | Supermicro X13SAE | 1 |
| CPU (a) | Intel Core i9-13900K 24C32T 3.0 GHZ 36 MB | 1 |
| RAM | Hynix 32 GB DDR5 4.8 GHz ECC Unbuffered | 2 |
| Storage 1 (b) | Western Digital SN850X 1TB NVMe | 1 |
| Storage 2 | Samsung U.2 PM9A3 3.84 TB PCI-E 2.5″ | 1 |
| GPU | Intel UHD 770 | 1 |
| Monitor | Philips C-LINE 279C9 27″ 4K | 1 |
| UPS (c) | Riello NPW 1500VA/900W<br>Riello NPW 1000VA/600W | 1 + 2 |

(a) item changed from i9-12900K to i9-13900K, please see subsection Total and Final Costs. (b) item changed from 500 GB to this 1 TB, please see subsection Total and Final Costs. (c) These UPS units were selected to support the four desktops, where two were planned to be supported by the single UPS unit with the highest capacity.

Based on Table 6, the cost for the four desktops amounted to EUR 16,161.46 (Supplementary Materials S6: Proposal S5), which, in average, resulted in EUR 4040.36 per unit.

As seen from this table, the scalability possibilities are essentially related to RAM capacity, frequency, and even GPU. Considering that the MB has a limit of 4.4 GHz and the modules enable 4.8 GHz, a possible replacement with a newer MB could bring about frequency improvements without modifying the RAM modules. Moreover, the CPU RAM specification enables a higher limit of 5.6 GHz. This further enables even more scalability by replacing the MB and RAM modules, enabling this performance to be achieved. Moreover, the system RAM capacity is limited to 192 GB, which can offer the following two ideal possibilities: (i) to acquire two identical modules and achieve 128 GB with a maximum of 4.8 GHz or (ii) to acquire four increased capacity modules of 48 GB, thus achieving a maximum of 192 GB and 5.6 GHz. In the desktops design, we did not consider a GPU due to the fact that the Intel CPU offers this integrated solution (Intel UHD 770), and in terms of data analyses, nothing was foreseen for machine learning or artificial intelligence [25], as these activities are in a development stage in our research team. However, at any moment, the design enables this possibility.

### 2.2.4. Workstation—The Higher-Throughput Computer

In addition to the desktops, we also designed a workstation providing extra capacity, which involved selecting server-like CPUs, such as the Intel Xeon. In opposition to the desktops, here, we could not find the DDR5 standard; hence, the design process became limited to the DDR4. Therefore, we planned the acquisition of a single workstation, leaving the possibility to design additional workstations at a time when funding is available and the industry can provide the DDR5 compatible CPUs, MBs, and RAM.

Hence, we selected the third-generation Intel Xeon CPUs, where we opted for the Xeon Gold 6348. This CPU provided the DDR4 3.2 GHz maximum frequency, which corresponded to the MB limits. Moreover, it provided ECC memory specifications, which have been available in this line of products for a longer time. Thus, we planned 12 modules of DDR4 ECC RAM, providing a total of 384 GB with a maximum frequency of 3.2 GHz. In this design, we conceived that double processing capacity could be provided in one system; thus, we considered a MB that could support two CPUs (also foreseen in the DC server). A design that has been successfully followed in previous works and research teams were the author has been involved originating nearly 13 Nexus workstations acquired over the years [29,30]. Contrarily to the previous desktop CPUs, the Xeon CPUs do not provide the graphics functionality. Hence, this design also required a GPU selected with minimum requirements.

In terms of storage, we selected three SSDs, where one with 500 GB would be used for the OS, one with 1 TB would be used for Linux/UNIX swap memory, and one with 4 TB would be used for data analyses. Additionally, with the possibility of performing Whole-Genome Sequencing (WGS)/WES analyses, including genome assembly, we also added a 4 TB HDD drive. This additional storage capacity could be applied, for instance, in storing temporary and intermediate files. Thus, the total combined storage capacity reached 9.5 TB, where the usable storage capacity (without the OS reserved space) can reach 8 TB. This design also considers a monitor and UPS (Table 7).

**Table 7.** Nexus workstation final design.

| Component | Specification | Units (Number) |
|---|---|---|
| Motherboard | Supermicro X12DAI-N6 (dual socket) | 1 |
| CPU | Intel Xeon Gold 5320 26C52T 2.2 GHz (52C104T) | 2 |
| RAM | Samsung 32 GB DDR4 3,2 MHz ECCR 2Rx4 (384 GB) | 12 |
| Storage 1 (*) | Western Digital SN850X 1 TB SSD NVMe | 2 |
| Storage 2 | Samsung U.2 PM9A3 3.84 TB SSD PCI-E 2.5″ | 1 |
| Storage 3 | HDD Seagate EXOS 7e10 4 TB SATA3 | 1 |
| GPU | PNY QUADRO T400 4 GB | 1 |
| Monitor | Philips 329P1H 31.5″ 4K | 1 |
| UPS | Riello NPW 2000VA/1200W | 1 |

(*) one unit changed from 500 GB to this 1 TB, please see subsection Total and Final Costs.

Based on Table 7, the estimated cost for the workstation reached EUR 11,978.97 (Supplementary Materials S6: Proposal S6). It is worth noting that the CPU Xeon Gold 6348 was modified in the process to this similar alternative, thus offering lower costs. In fact, the differences from the 6348 and the 5320 essentially lie in the following three aspects: (i) in the number of cores, thus being 2C4T, (ii) in the cache memory size, thus being 3 MB, and (iii) in the RAM frequency, thus being 267 MHz. We recognize that this can constitute a trade-off when these small differences can be overcome with a higher difference in cost. By modifying the CPU, we reduced the total cost to an estimated amount of EUR 4000 to EUR 4500 for both CPU units.

In terms of scalability, this design enables an increase in RAM capacity, storage, and GPU. It offers a RAM limit of 4 TB, which provides plenty of capacity for scalability. The enclosure provides room for several disk drives, thus enabling a storage capacity increase. Moreover, the current GPU could be used in parallel or replaced with a better GPU capable of performing machine learning or artificial intelligence model training analyses and molecular dynamics analyses [25].

### 2.2.5. Network and Connectivity

Beyond the novel computational designs, this opportunity was also used to acquire additional network equipment, thus not only enabling access to the servers, but also to complement the room adaptation presented in the previous subsection—Existing Resources and Adaptations. These resources are listed in Table 8.

Without considering the final power strips (Table 8 (b); Supplementary Materials S6: Proposals S1 and S4), the total cost was EUR 2484.75, where the fiber-optics cable was acquired at a different time with a cost of EUR 124.29 (Supplementary Materials S3) and the fiber-optics adapters were acquired as part of the budget in Table 2 (Supplementary Materials S4), both from the CICS-UBI funding. During acquisition, the Nexus company introduced the final power strips, which came to replace the smaller ones (Table 8 (a), “not final”), increasing their cost (Supplementary Materials S6: Proposal S4 and Supplementary Materials S2).

In terms of scalability, we can only refer to the novel managed switch, which, after the room adaptation, leaves the possibility to activate ten additional RJ45 1GBE ports and also the possibility to install a second fiber-optics cable to be connected in its second free Small Form-Factor Pluggable (SFP)+ port. This is also congruent with the availability of the second free RJ45 10GBE port on the DC server, for which we could use the second copper cable (Table 8). These two 10 GBE additional ports became available with the inclusion of a BASE-T Network Interface Card (NIC; Table 4), thus supporting high-speed connections.

The secondary fiber-optics cable would double the speed and capacity of data transfers to and from the DC server and also increase the number of available connections beyond the mentioned ten with the acquisition of an additional similar managed switch (which could be installed in the same rack). No wireless access was considered in this design, as all the computers, new and old (including servers), provide from one to two RJ45 1GBE connections and the novel ones do not provide wireless support (Tables 4–7). We further gave preference to wired connections, as these typically do not require credentials, thus facilitating access. Wireless access is still available through the existing networks, but these are not granted to access the bioinformatics planned Virtual Local Area Networks (VLANs) and the servers (Supplementary Materials S1: Figure S3).

**Table 8.** Network and connectivity equipment final design.

| Component | Specification | Units (Number) | Cost (EUR) |
|---|---|---|---|
| Managed switch (a) | Netgear M4300-28G GSM4328S-100NES 24X10GBE + 4X10GBE | 1 | 1519.05 |
| Switch (b) | TP-LINK SG-105 1GBE | 3 | 67.16 |
| Copper Cables (a) | Fscom cat8—1.5m | 2 | 49.20 |
| GBIC (a) | Netgear Prosafe 10G SR SFP + Multi Mode<br>Fscom 10GBE SFP+ > 10GBE Base-T<br>Fscom 10GBE SFP+ | 1 + 1 + 1 | 455.10<br>227.55<br>47.97 |
| Power strips (not final) (a) | Nexus Energy Ruler 5 Outlets White Without Switch | 9 | 46.38 |
| Power strips (final) (b) | Monolyth 3050001 19″ 8 sockets with switch and 16A Schuko cable | 9 | 203.02 |
| Fiber optics cable (c) | Patch cord FO LC-LC duplex OM3 70 m | 1 | 124.29 |
| Fiber optics adapters (d) | Coupler LC/LC multi-mode OM3<br>Cable fiber optics LC/SC 1m multi-mode OM3 | 1 + 1 | 15.21 |

(a) material initially listed regarding October 2022 (Supplementary Materials S6: Proposals S1 and S4); (b) changes with their value as part of the final cost from March 2023 shown in Table 9 (Supplementary Materials S2: Proposal S2); (c) acquired separately from CICS-UBI funding (Supplementary Materials S3); (d) acquired together with the Table 2 set of components (Supplementary Materials S4) from CICS-UBI funding.

This initial plan focused on ensuring the cybersecurity and performance of the networks by implementing two VLANs (1 and 2) by considering the restrained access of users and devices [55]. VLAN1 attempts to encompass all researchers, devices, and spaces requiring bioinformatics access, whereas VLAN2 is only focused on encompassing the servers dedicated to genomic medicine. This plan attempts to ensure the following four conditions: (i) access to the servers located in VLAN2 is only attained by researchers using or registered to VLAN1 and also by those possessing the necessary server credentials; (ii) access to VLAN1 is only attained upon previous registration of the personal or institutional computer/laptop (MAC address) or by adding the new space room/office to the configuration; (iii) other offices and spaces can, in future, be added or removed, thus gaining more control over user access; and (iv) VPN is available to enable remote access but only to VLAN1, from where it can be possible to access VLAN2.

### *2.3. Total and Final Costs*

Here, we present the costs per acquisition (Table 9), thus considering the four purchase moments without considering the chronological order, (i) novel computational resources, (ii) upgrades to the existing desktops, (iii) the dry-lab adaptation material, and (iv) additional connectivity.

**Table 9.** Total costs per acquisition and total amount.

| Category | Cost (EUR) | Year | Company | Reference Material |
|---|---|---|---|---|
| Novel computational resources | 51,482.19 | 2023 | Soon—Business Solutions, S.A./Nexus Solutions, S.A., Maia, Portugal | Tables 4–7, Table 8 minus (a,c,d) and Supplementary Materials S2: Proposal S2. |
| Upgrades to existing desktops | 3468.63 | 2024 | RuiPolana/Eurobit—Sistemas Informaticos e Manutenção, Lda, Covilhã, Portugal | Tables 2 and 8 (d) and Supplementary Materials S4. |
| Dry lab adaptation material | 3508.14 | 2023 | L3W Material Eléctrico, Lda, Famalicão, Portugal | Supplementary Materials S5. |
| Additional connectivity (fiber cable) | 124.29 | 2023 | Copper2Fiber—Soluções de Conectividade, Lda, Barcarena, Portugal | Table 8 (c) and Supplementary Materials S3. |
| Total | 58,583.25 | - | - | - |

By summing the partial costs from October 2022 described in the previous subsection—Novel Computational Resources—which do not consider the nine power strips (EUR 46.38), three UPS units (EUR 616.23), and three mini switches (EUR 67.16), the total cost reached EUR 54,134.66 (Supplementary Materials S6). However, during acquisition later in 2023, whilst considering all planned material and after a few updates due to out-of-stock components (while maintaining the designs), this value initially dropped to EUR 53,266.38 (Supplementary Materials S2: Proposal S1) and later after a renewal of the proposal validity to the value listed in Table 9 (Supplementary Materials S2: Proposal S2). This was likely due to the particular inflation variation felt in this period. It is clear that the final designs previously presented (Tables 4–8) show some differences from the final proposals (Supplementary Materials S2). These occurred most likely due to stock changes of the Western Digital (WD) SN850 SSDs of 500 GB and 1 TB, thus being replaced by WD SN850X 1TB SSDs by the Nexus company. Regarding the room adaptation material, it is clear that the value is lower than that indicated in the proposal (Supplementary Materials S5), which is due to unacquired material. Similarly, the upgrades to existing desktops (Supplementary Materials S4) only considered the selected material (Table 9). Hence, the total expenditure of the whole intervention amounted to EUR 58,583.25, spanning the years 2023 and 2024.

## 3. Discussion and Evaluation

### 3.1. General Remarks

Here, we report the work developed for the design, implementation of informatics equipment and resources adaptation for the establishment of a bioinformatics laboratory in CICS-UBI. This comprised a dedicated bioinformatics common space (i.e., dry-lab room) and novel equipment specifically prepared for our research team. Although in different settings, a few works were found to describe the development of resources and infrastructures for bioinformatics (e.g., [6,13,18,23,31,48–52]), to our knowledge, we have not found any other similar and comprehensive work published. As previously mentioned, we established our initial bioinformatics resources, which were developed while using funds, not sufficient for establishing a cloud computing or HPC high-end computational resources (e.g., [18,48]). Nevertheless, the workstation and servers designed (Tables 4, 5 and 7) can be regarded as a first step towards implementing a higher scale HPC resource. While in future acquiring many similar workstations and/or server nodes, it becomes possible to transform them into a computing cluster or possibly a (private) cloud resource, for instance, using Hadoop software (e.g., [49]). In this sense, this work can be regarded as a base for establishing and comparing to such high-end developments both in terms of capabilities, performance, and costs. On the other hand, it is understandable that cloud or other HPC resources are currently regarded as the optimal options for processing big data [48,75] while also seeking to enforce data protection [75,76]. However, our work is focused on providing local resources and infrastructures, where the advantage stems mostly from protecting sensitive patient genomics data [75–77]. In this sense, we opted to minimize the risk by keeping such sensitive data within the university boundaries, thus avoiding the use of external or foreign computing resources (e.g., by metadata exposure, human error, among others [75,76]).

The dry-lab room was adapted to ensure network and electrical connections for all desktops and even additional computers, such as laptops potentially brought in by students and researchers. Due to biodata sensitivity [13,24,76,77] and ethical matters, we additionally foresaw a plan for the establishment of novel VLANs (Supplementary Materials S1: Figure S3). Beyond this novel equipment, we also introduced upgrades to the existing five desktops (Table 3), thus improving the capacity and performance of the dry-lab computational resources.

For our team, we developed four designs, two for servers, DC (Table 4), and DCB (Table 5), one for desktops (Table 6), and one for workstations (Table 7). These were developed in order to provide state-of-the-art hardware, thus ensuring that they provide the highest performance possible. However, this objective can arise as a utopia, since it is often dependent on the budget dispensed and the availability of the optimal equipment on the market [53,54]. In fact, as companies provide different stocks and may have different suppliers from different brands and considering the novel standards, the options available are usually limited [53,54]. This is also true for companies such as Nexus. However, it is recognizable that the scalability possibilities seem to be better regarding the Nexus material when compared to the material from Table 1 [53,54]. For instance, the HP desktops CPUs could, in principle, offer an upgrade to 128 GB RAM, but according to the supplied specifications, this is limited to 64 GB.

Three Nexus desktops had the aim of being used by single members of our team, and, hence, they were planned to be located in the two researchers' offices, whereas the fourth one would be dedicated for common use in the wet-lab. In case the desktop's processing capacity was not sufficient for certain analyses, we designed a workstation. This workstation can be regarded as a (scalable) mini computing cluster, which, when combined with the DC, can complement it with a higher processing capacity (e.g., [31]). Thus, offering a more automated and independent data processing capability, which could be relevant, for instance, in accessing local bioinformatics platforms for use by various researchers [31].

Regarding the HDD storage choices, we opted for Serial Advanced Technology Attachment (SATA) drives as opposed to the SAS [78], and to ensure that they could provide 4Kn AF [72–74]. The choice of SATA considered that this technology can also be applied to the server context through opting for a line of products especially designed for the enterprise market and still withstanding lower costs when compared to SAS alternatives. In fact, considering the amount of data to be stored and the contained utilization by our team, we decided that the SATA technology, despite having a few disadvantages (i.e., less speed, security, among others) [78], would be sufficient for our needs.

In all designs, we considered SSDs for booting following the (SATA/Non-Volatile Memory express—NVMe) technology, which provide a better performance. Only in the case of the workstation and both servers we additionally opted for SATA HDDs. The HDD in the workstation was selected to function as a temporary data backup drive during data analysis, where many temporary files can be stored. These mechanical and enterprise-level devices are nowadays considered more resilient [78]. When compared to SSDs, they provide far less speed (e.g., ~250 MB/s—SATA HDD vs. ~550 MB/s—SATA SSD vs. ~7000 MB/s—NVMe SSD) and are more prone to failures. However, for the same capacity, HDDs currently present a lower cost (nearly half), thus facilitating server designs and replacements. In many SSD and HDD cases, it is possible to find 4Kn sector sizes, a format that was intended to be applied and available in all designs. However, some of the SATA/NVMe SSDs, namely those acquired for booting the servers and for upgrading the G.Skill desktop (herein exception made to the WD NVMe SSD), surprisingly, do not seem to enable the 4Kn format. In fact, this information is not easy to find associated with the material, even upon consulting the companies. Hence, it becomes difficult to ensure all the planned material supports this format. Contrarily, during design, we made an informed decision regarding the Seagate SATA HDDs and other NVMe SSDs found in the servers, workstation, and desktops.

Another matter of research concerns the selection of drive models comprising the server's pools [79]. To our knowledge, these can be found assembled using a single model, thus creating the potential risk of any manufacturing defects affecting pools nearly at the same time [79]. Hence, we decided to make an attempt to avoid such a scenario, and two different models and capacities (Tables 4 and 5) were selected, albeit from the same brand. Thus, each pool consists of half Seagate Exos 7e8 and half Exos 7e10, which offer very similar performances. Similarly, this option was also applied to the DC boot pool (Table 4). Even though, considering that defective drives could originate from the same batches or company [79], they could still fail at approximate times and affect all pools. However, it is also true that our pools were designed considering different applications, drive capacities/models, and usages; hence, the likelihood of failure of several pools at near times was considered to be lower.

### 3.2. Computational Resources and Bioinformatics Applications

Here, we make an attempt to highlight the potential applications of the presented designs to some of the bioinformatics areas of study, and in another sense, the potential bioinformatics analysis our system is intended to support, coupled with a more detailed comparison of all the designs.

The design of computational resources for bioinformatics and computational biology is not an accurate and objective work. The designs and configurations often depend on the bioinformatics field of work being conducted, the software tools being applied, developed and of the amount and complexity of data involved (e.g., [23]). The computational resources will naturally get wore off, outdated, and obsolete, often requiring upgrades, or, in the case that these are not possible, their replacement. Hence, in a bioinformatics laboratory, the acquisition of novel resources is often required. This becomes necessary not only to replace the existing ones (e.g., end-of-life), but also to enable additional capabilities, thus allowing for continued formation and researchers productivity. In this sense, it is also true that computers presently acquired will be able to handle multiple analyses and software development, but within some time, other specifications might be needed to accompany the scientific field. This can be analogous to acquiring wet-lab scientific equipment and consumables. In a few years, the desktops provided with upgrades (Table 3) are expected to be the first requiring replacement. It would be ideal to replace these desktops with newer workstations providing data integrity and improved computing performance combined with ECC and 4Kn specifications. With the help of private external storage, these desktops were initially capable (Table 1) of handling some microbial genome analyses (e.g., [46]), but still were not provided with sufficient RAM capacity and storage to handle some more demanding bioinformatics tools, such as those from genomic medicine [10,17] here performing exome sequence analysis or, for instance, metagenomic analyses [80,81].

Moreover, they can now handle multiple users, software development, bigger/multiple data sets, downstream data analyses, and several bioinformatics tools. Hence, they are in a better position, but it is not possible to ensure that they will be capable of handling any demanding bioinformatics tasks or analyses. Despite the novel desktops (Table 6) and workstation (Table 7) are more recent with better specifications, this also applies. The novel acquisitions provide comparatively increased competencies (Table 10) in terms of CPU (number of cores, ECC, frequency, among others) and RAM (DDR5, capacity, and frequency). In our experience, these desktops have been capable of handling the complex and large spreadsheet documents typically employed for the final steps of WES data interpretation [17]. Specifically, they are capable of opening large spreadsheet documents (~2 GB) instantly, which was previously nearly impossible. Our new laboratory will have application not only in research, projects, and data analyses, but also contribute to the formation of students in these areas. As such, the number of bioinformatics software used, installation, and utilization can be very diverse, considering the different omics data and respective analyses that can be handled [41,75]. For instance, we are currently using software such as Seaview v5.05 [82,83], IGV v2.16 [84], SpliceAI v1.3.1 [85], PAML v4.9 [86], PhyML v3.3 [87], BLAST v2.12 [88], Bioconda [89], SAMTools v1.19.2 [90], BAMtools v2.5.2 [91], SRA Toolkit v3.2.1 [92,93], FastQC v0.12 [94], FLASH v.1.2.11 [95], Trim Galore! v0.65 [96], SPAdes v3.15.5 [97], QUAST v4.0 [98], Busco v5.5.0 [99], Ragout v2.3 [100], and NOVOPlasty v4.3.3 [101], among others.

**Table 10.** Detailed comparison of the computational resources.

| **System** | **HS** | **Desktop** | | | **Workstation** | **Server** | |
|---|---|---|---|---|---|---|---|
| Brand | na | HP | G.Skill | Nexus | Nexus | Nexus DCB | Nexus DC |
| OS | na | Kubuntu 24.04 LTS | | | | TrueNAS SCALE 24.04.2 | |
| Kernel | na | 6.8.0-41-generic #41-Ubuntu | | | | 6.6.32-production+truenas #1 | |
| CPU | Cores/Threads | 6C12T | 8C16T | 24C32T | 52C104T (2x 26CT52) | 6C12T | 16C32T |
| | Clock Frequency (GHz) | 3.2–4.6 | 3.7–4.35 | 2.2–5.8 | 2.2–3.4 | 2.9–4.8 | 2.9–3.5 |
| | Cache (MB) | 12 | 16 | 36 | 78 (2x 39) | 12 | 24 |
| RAM | Standard | DDR4 | DDR4 | DDR5 | DDR4 | DDR4 | DDR4 |
| | Clock Frequency (GHz; installed) | 2.6 | 2.4 | 4.8 | 3.2 | 3.2 | 3.2 |
| | Clock Frequency (GHz; working) | 2.6 | 2.4 | 4.4 | 2.9 | 3.2 | 3.2 |
| | Size (GB) | 64 | 64 | 64 | 384 (2x 192) | 32 | 128 |
| ECC | Supported | No | No | Yes | Yes | Yes | Yes |
| Storage | Usable SSD | 4 TB | 4 TB | 3.84 TB | 3.84 TB | na | na |
| | Usable HDD | na | 2 TB | na | 4 TB | 48 TB | 48 TB + 80 TB |
| | Total (*) | 4 TB + 256 GB (#) | 6 TB + 2 TB (#,$) | 3.84 TB + 1 TB (#) | 7.84 TB + 2 TB (#,$) | 48 TB + 256 GB ($) | 128 TB + 480 GB ($) |
| | 4Kn Supported Drives | All | Usable SSD | All | All | Usable HDD | Usable HDD |

(*) Total: beyond usable drives capacity, also shows the additional OS dedicated drives. Following attached symbols indicate: (#) drive reserved for swap memory and ($) drive(s) reserved for OS installation/boot. na—not applicable. HP: Hewlett-Packard. HS: Hardware Specification.

Although, the workstation provides six times more RAM capacity and 3.25 times more CPU processing cores/threads, it is not provided with the DDR5 memory standard or having the highest CPU clock frequency. However, this aspect could be compensated for by the higher number of cores/threads. With some exceptions, the Xeon CPUs have followed this trend, comparatively offering many cores and a lower frequency. The cache memory is the highest in the workstation, which contributes to improve the performance of the

CPU. Hence, the difference in RAM could be the only weakness potentially limiting its performance when compared with the Nexus desktops (Table 11).

The workstation was designed to be useful during research activities when more capacity is required. It was planned to be a shared common computer where several researchers may perform their data analyses at the same time by using a fraction of its processing capacity (e.g., Multi-core Method for Analysis Pipelines (MMAP) [28–30]) and without occupying a permanent seat. Additionally, the workstation was also designed by having in mind the possibility of performing more demanding NGS data analyses (e.g., assembly, annotation, among others), which can reveal to be of great importance for assessing the complete human WGS composition or for performing metagenomics analyses [80,81]. To help with the data processing in these resources, we also planned the servers, enabling the centralized storage, retrieval, and analyses of WGS/WES data at any moment. This was achieved by also planning a more capable NIC (Table 4) and related devices (Table 8) to be useful in cases where the server needs (i) to connect several members of the team, (ii) to reach UBI-CIS computational resources for data backups, or (iii) even reach the internet for NGS data transfers, which can be received through the cloud services, such as AWS.

### *3.3. Computational Resources Performance Evaluation*

In order to provide more objective results regarding the systems described, we sought to evaluate the performance on the novel designs (Tables 4–7) and upgraded resources (Table 3). In this sense, the works from Boras et al. [102] and Balen et al. [103] could be adapted to our purpose. Our objective here is to provide an overall indication where the hardware differences and systems performances are evaluated while employing the same OS (Table 10). From the various utilities available (e.g., [102,104–110]), we found that UnixBench v.5.1.3 [103,111] is a more practical benchmarking solution, suitable not only for Unix-like systems, but also for an overall system benchmark (not specifically targeted for CPU, RAM, among others) [111]. This software was especially used for analyzing the performance of Virtual Private Servers [103], which could also be adapted not only for our bare metal servers, but also for analyzing the performance of our desktops and workstation. By using the same benchmarking utility, this could also bring comparable results among all systems designs, regardless of the design at hand. Nevertheless, this comparison can be biased by the different OS (Table 10).

Based on the works of Boras et al. [102] and mainly of Balen et al. [103], we devised simple methodologies and conditions for the benchmark tests. In the case of non-server system resources, we established routine 1, as found in Supplementary Materials S7: File S1.

In the case of server systems, the previous routine required a few adaptations, thus establishing routine 2, as found in Supplementary Materials S7: File S1.

The results from the two separate routines are presented in Table 11 for each system design.

From this set of results, it is possible to understand the slight differences among the three UnixBench runs and modes in each system case, which attests for the similar running conditions. Further detailed results and tests undertaken can be found in the Supplementary Materials S7: File S2. Overall and as expected, the (average) results show that the novel computational resources (Tables 4–7) perform better than the previously existing and upgraded resources (Table 3). Of note is the Nexus desktop, which presents the highest result for the same OS, only defeated by the Nexus DC Multi-core results with a different OS (see alto Table 10). These results demonstrate that this desktop design has the highest performance. On the other hand, it was not expected that the Nexus workstation would achieve lower results, mostly considering the highest number of cores and RAM capacity. Additionally, this system presents lower clock frequencies in both RAM and CPU. In terms of server benchmarking, it is surprising that the DCB shows the highest results in terms of the Single-core mode and worse results when in the Multi-core mode, when compared to the DC server, which performs better here and worse in the Single-core mode.

**Table 11.** UnixBench benchmarking of complete systems designs described in Table 10.

| Benchmark (*) | | Desktop | | | Workstation | Server | |
|---|---|---|---|---|---|---|---|
| **Run** | **Mode** | **HP** | **G.Skill** | **Nexus** | **Nexus** | **Nexus DCB** | **Nexus DC** |
| 1 | Single-core | 999.6 | 1058.4 | 1920.4 | 1534.5 | 1666.6 | 1214.1 |
| | Multi-core | 5935.7 | 6358.5 | 15755.8 | 12784.6 | 9424.0 | 17088.8 |
| 2 | Single-core | 966.0 | 1054.8 | 1923.0 | 1535.0 | 1676.2 | 1206.3 |
| | Multi-core | 5947.7 | 6371.6 | 15822.5 | 12360.5 | 9445.1 | 17174.3 |
| 3 | Single-core | 979.4 | 1053.8 | 1919.2 | 1525.9 | 1666.0 | 1195.8 |
| | Multi-core | 5947.7 | 6371.0 | 15870.8 | 12655.6 | 9449.0 | 17029.0 |
| Average | Single-core | 981.7 | 1055.7 | 1920.9 | 1531.8 | 1669.6 | 1205.4 |
| | Multi-core | 5943.7 | 6367.0 | 15816.4 | 12600.2 | 9439.4 | 17097.4 |

(*) all values are gathered from the given UnixBench "System Benchmarks Index Score".

All the UnixBench tests and runs were performed as offered by default and without testing the GPU devices, which are not currently our focus. However, as a future work, this and additional benchmarking platforms and tests could be worth presenting more thoroughly. This would provide an assessment of our systems standpoint and highlight aspects for improvement.

### *3.4. Comparison with EMBL-EBI Course Material from 2024*

Although our implementation can be very different from a more advanced "Bioinformatics Core Facility" (BCF), here, we dedicate some space to highlighting the current differences from both cases.

In the following Table 12, we reintroduce the main topics found in the EMBL-EBI course material from 2024 and add a third column to show how our current implementation compares to the existing core facility development areas.

From Table 12, the great differences found from our current implementation standpoint to what can already be found developed in major BCFs like those from EMBL-EBI are clear. Our current resources and infrastructures cannot currently be regarded as a BCF, as we lack tackling many of the areas shown in Table 12. In fact, our novel resources were conceived to be limited to the use of our team and the upgraded resources available in dry-lab for the common use of our research center by also other teams. Knowing that these are only regarding initial conditions, to meet the BCF conditions, we would require additional computational and (stable) human resources for the majority of the areas, especially those proficient in Bioinformatics, Computer Sciences, Mathematics, and in diverse omics, who among others, can offer training and brokering, develop new resources, and research new bioinformatics methods and solutions. Here, we also propose to include two recent areas of increasing attention and development, namely, the Cybersecurity, respecting the implementation of measures of avoiding intrusions and cyberattacks, and the Data sensitivity, privacy and protection, which respect the development of novel bioinformatics strategies to make the identification of patients more difficult. This can be possible through their genomic information and associated metadata, hence, they are currently receiving more attention from the scientific community, as this would avoid disruption to human patient ethics.

**Table 12.** Comparison of our current standpoint to the EMBL-EBI BCF development areas.

| Development Areas | 2003 | 2020 | This Work |
|---|---|---|---|
| Hardware resources (servers, cluster) | Yes | Yes | Yes * |
| Local copies of public databases | Yes | No | No * |
| Disk storage | Yes | Yes | Yes |
| Help-desk via email | Yes | Yes | Yes |
| Web-site (help documentation) | Yes | Yes | No * |
| Web-site (web-based programs) | Yes | No | No * |
| One-to-one tutorial sessions | Yes | Yes | No * |
| Formal teaching (short practical training courses) | Yes | Yes | No * |
| Formal teaching (undergraduate/masters teaching) | Yes | Yes | Yes * |
| Development of scripts, pipelines, and interfaces—research | Yes | Yes | Yes * |
| Project-based consultation and collaboration | Yes | Yes | Yes * |
| Brokering, skills sharing, advocacy | Yes | Yes | No * |
| Grant writing | No | Yes | Yes |
| Project-specific databases | Yes | Yes | No * |
| Sample-tracking/LIMS development | No | Yes | No |
| Analysis as a service | Yes | Yes | No * |
| Support for other Bioinformatics and core facilities | No | Yes | No * |
| Web hosting | No | Yes | No * |
| Tissue banking data infrastructure—biobank | No | Yes | Yes * |
| Cybersecurity ** | - | - | Yes * |
| Data sensitivity, privacy and protection ** | - | - | No * |

* Potential cases for future work or in development. ** Novel areas from our case.- no information.

## 4. Conclusions

Here, we presented work developed over two years to design and adapt informatics equipment to the specific conditions in our building rooms and offices at the CICS-UBI. This work provided many new insights and learning into the novel technologies and how to make better choices and adapt them to the current needs. This project enabled knowledge updates and learning in many novel aspects, among others related to RAID technologies, network, hardware, NAS operating systems, HDD and SSD metrics, file systems, benchmarking, and including experience in services management.

Additionally, in all designs, the important aspect of scalability was considered. Here, the planning based on the current conditions and available information can change in the future. Thus, it is always good practice to provide room for changes and adaptations to novel situations, mainly while working with limited funding. We thus designed desktops (four identical), one workstation, and two servers together, with extra components for storage scalability and network connectivity. All novel designs provide room for additional storage, GPU adaptations, RAM, storage, among other possibilities. The bioinformatics laboratory (or dry-lab) was also target of technical transformations, thus incorporating the novel managed switch and the necessary cabling within the room (Ethernet/RJ45 cables and power cables) and connecting the room (fiber cable). All designs provide increased capacity for bioinformatics, as shown by the performance evaluations. In the future, we hope to perform acquisitions of additional spare components, such as HDDs, UPS, and other material. Moreover, we initially decided that the DC server NGS data backup was not necessary and was left for external storage drives. However, in case that this volume becomes unavailable (e.g., due to any malfunctions), the time period to recover from this unavailability could become greater than expected, thus affecting the progress of bioinformatics activities. Hence, in future, we could consider the acquisition of a second backup server for this type of data. However, this hypothetical novel acquisition does not remove the importance of maintaining the external backup drives. Furthermore, considering that companies have been removing Ethernet ports from their computer products (e.g., laptops), another future improvement would be to acquire wireless equipment to extend the initial network plan and facilitate connectivity to bioinformatics infrastructures. Any of these cases will be important to ensure that any unforeseen changes or repairs can be immediately handled, thus avoiding longer periods of resolution time and possibly limited funds. Moreover, the acquisition of GPU components or novel designs of desktops/workstations/servers with more recent GPUs will be especially important for the aforementioned bioinformatics applications, thus expanding the potential of bioinformatics research activities in CICS-UBI. We hope that this work published as a case study can inspire and contribute as an example to other researchers seeking to implement their bioinformatics resources or laboratories.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/technologies13070285/s1, Supplementary Materials S1: Relative rooms location and adaptation plans. Supplementary Figure S1: Illustration of available rooms and relative location. Supplementary Figure S2: Dry-lab infrastructure plan and network connectivity. Supplementary Figure S3: Network configuration plan considering all rooms and spaces. Supplementary Materials S2: Novel computational resources designs updated showing total costs during acquisition. Supplementary Proposal S1: All designs with total cost in January 2023 (186). Supplementary Proposal S2: All designs with final total cost in March 2023 (2093). Supplementary Materials S3: Dry-lab network connectivity. Supplementary Proposal S1: Fiber optics cable with ready tips (524_2023). Supplementary Materials S4: Desktop's hardware upgrades. Supplementary Proposal S1: List of alternative hardware components available for upgrades (Table 2) including the fiber optics adapters (Table 8). Supplementary Materials S5: Dry-lab room adaptation. Supplementary Proposal S1: Dry-lab room adaption technical material (40/3363). Supplementary Materials S6: Novel computational resources finalized designs dated from October 2022 showing approximate partial costs. Supplementary Proposal S1: DC server design with network and connectivity material (6998). Supplementary Proposal S2: DCB server design (7003). Supplementary Proposal S3: Extra HDD drives for DC and DCB servers (7004). Supplementary Proposal S4: Single power strip with five outlets (to be multiplied by nine) (7005). Supplementary Proposal S5: Desktop design (7007). Supplementary Proposal S6: Workstation design (7008). Supplementary Materials S7: Benchmark routines and detailed information. Supplementary File S1: Benchmark routines for systems designs performance assessment. Supplementary File S2: Benchmark runs and tests detailed results for each system design.

**Author Contributions:** Conceptualization, E.M. and M.C.L.; methodology, E.M.; software, E.M.; validation, E.M. and M.C.L.; formal analysis, E.M.; investigation, E.M.; resources, E.M. and M.C.L.; data curation, E.M.; writing—original draft preparation, E.M.; writing—review and editing, E.M. and M.C.L.; visualization, E.M.; supervision, E.M. and M.C.L.; project administration, E.M. and M.C.L.; funding acquisition, M.C.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Portuguese Foundation for Science and Technology (FCT) (grant UIDB/00709/2020 (https://doi.org/10.54499/UIDB/00709/2020)), and by Programa Operacional Regional do Centro (grant CENTRO-08-5864-FSE-000039). The funding bodies did not play any roles in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in this study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding author.

**Acknowledgments:** We would like to thank to Paulo Gouveia and Rui Sousa from the company Nexus Solutions, S.A., for all the advice and assistance given while designing the solutions; to Rui Costa from the CICS-LTIS, for all the advice given while preparing the servers and network connectivity considering the UBI infrastructures and facilities; to the members of the TrueNAS community forum for the helpful comments; to Vitor Soares and team from the UBI-TIS for the helpful suggestions and assistance regarding the provisioning of rooms adaptation; to José Barata from Eurobit—Sistemas Informáticos e Manutenção, Lda in association with RuiPolana, for all the patience, preparedness and kind assistance with budgets and equipment enabling work advances; and to Gabriela Rocha from the company Divultec, Serviços de Informática Lda for the assistance in establishing the possibilities for the hardware upgrades. We would like to thank the editorial team and three anonymous reviewers for the helpful comments and suggestions which helped to improve a first version of the manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

| | |
|---|---|
| 4Kn | 4KB native sector size |
| AF | Advanced Format |
| AWS | Amazon Web Services |
| BCF | Bioinformatics Core Facility |
| C4-UBI | Cloud Computing Competence Centre |
| CICS-UBI | Health Sciences Research Centre |
| CICS-LTIS | Local Technical Informatics Services |
| CPU | Central Processing Unit |
| DC | Data Center server |
| DCB | Data Center Backup server |
| DDR | Double Data Rate |
| dRAID | declustered RAID |
| ECC | Error Correction Code |
| EMBL-EBI | European Molecular Biology Institute-European Bioinformatics Institute |
| GBE | Gigabit Ethernet |
| GBIC | Gigabit Interface Converter |
| GPU | Graphical Processing Unit |
| HBA | Host Bus Adapter controller |
| HDD | Hard Disk Drive |
| HP | Hewlett-Packard |
| HPC | High-Performance Computing |
| HS | Hardware Specification |
| HTC | High-Throughput Computing |
| IPMI | Intelligent Platform Management Interface |
| JBOD | Just-a-Bunch-of-Disks |
| LIMS | Laboratory Information Management System |
| MAC | Medium Access Control address |
| MB | Motherboard |
| MMAP | Multi-core Method for Analysis Pipelines |
| NAS | Network-Attached Storage |
| NGS | Next-Generation Sequencing |
| NIC | Network Interface Card |

| | |
|---|---|
| NVMe | Non-Volatile Memory express |
| OpenZFS | Open Zettabyte File System |
| OS | Operating system |
| PAS | Private Accounting Storage |
| RAID | Redundant Array of Independent Disks |
| RAM | Random Access Memory |
| RJ45 | Registered Jack-45 |
| SAS | Serial Attached SCSI |
| SATA | Serial Advanced Technology Attachment |
| SFP+ | Small Form-Factor Pluggable + |
| SSD | Solid-State Drive |
| UBI | University of Beira Interior |
| UBI-CIS | Coordinating Informatics Services |
| UBI-TIS | Technical Infrastructures Services |
| UHD | Ultra-High Definition |
| UPS | Uninterruptible Power Supply |
| VLAN | Virtual Local Area Network |
| WD | Western Digital |
| WES | Whole-Exome Sequencing |
| WGS | Whole-Genome Sequencing |

## References

1. Hagen, J.B. The Origins of Bioinformatics. *Nat. Rev. Genet.* **2000**, *1*, 231–236. [CrossRef] [PubMed]
2. Hogeweg, P. The Roots of Bioinformatics in Theoretical Biology. *PLoS Comput. Biol.* **2011**, *7*, e1002021. [CrossRef] [PubMed]
3. Gómez-López, G.; Dopazo, J.; Cigudosa, J.C.; Valencia, A.; Al-Shahrour, F. Precision Medicine Needs Pioneering Clinical Bioinformaticians. *Brief. Bioinform.* **2019**, *20*, 752–766. [CrossRef] [PubMed]
4. Lewitter, F.; Rebhan, M.; Richter, B.; Sexton, D. The Need for Centralization of Computational Biology Resources. *PLOS Comput. Biol.* **2009**, *5*, e1000372. [CrossRef]
5. University of Beira Interior CICS-UBI Health Sciences Research Center, University of Beira Interior. Available online: https://www.ubi.pt/sites/cics/en (accessed on 13 June 2024).
6. Courneya, J.-P.; Mayo, A. High-Performance Computing Service for Bioinformatics and Data Science. *J. Med. Libr. Assoc.* **2018**, *106*, 494–495. [CrossRef]
7. Regateiro, F.J.; Silva, H.; Lemos, M.C.; Moura, G.; Torres, P.; Pereira, A.D.; Dias, L.; Ferreira, P.L.; Amaral, S.; Santos, M.A.S. Promoting Advanced Medical Services in the Framework of 3PM—A Proof-of-Concept by the "Centro" Region of Portugal. *EPMA J.* **2024**, *15*, 135–148. [CrossRef]
8. Alhajaj, K.E.; Moonesar, I.A. The Power of Big Data Mining to Improve the Health Care System in the United Arab Emirates. *J. Big Data* **2023**, *10*, 12. [CrossRef]
9. Kong, D.; Yu, H.; Sim, X.; White, K.; Tai, E.S.; Wenk, M.; Teo, A.K.K. Multidisciplinary Effort to Drive Precision-Medicine for the Future. *Front. Digit. Health* **2022**, *4*, 845405. [CrossRef]
10. Petersen, B.-S.; Fredrich, B.; Hoeppner, M.P.; Ellinghaus, D.; Franke, A. Opportunities and Challenges of Whole-Genome and -Exome Sequencing. *BMC Genet.* **2017**, *18*, 14. [CrossRef]
11. Beckmann, J.S.; Lew, D. Reconciling Evidence-Based Medicine and Precision Medicine in the Era of Big Data: Challenges and Opportunities. *Genome Med.* **2016**, *8*, 134. [CrossRef]
12. Alyass, A.; Turcotte, M.; Meyre, D. From Big Data Analysis to Personalized Medicine for All: Challenges and Opportunities. *BMC Med. Genom.* **2015**, *8*, 33. [CrossRef] [PubMed]
13. Mulder, N.J.; Adebiyi, E.; Adebiyi, M.; Adeyemi, S.; Ahmed, A.; Ahmed, R.; Akanle, B.; Alibi, M.; Armstrong, D.L.; Aron, S.; et al. Development of Bioinformatics Infrastructure for Genomics Research. *Glob. Heart* **2017**, *12*, 91–98. [CrossRef] [PubMed]
14. da Fonseca, R.R.; Albrechtsen, A.; Themudo, G.E.; Ramos-Madrigal, J.; Sibbesen, J.A.; Maretty, L.; Zepeda-Mendoza, M.L.; Campos, P.F.; Heller, R.; Pereira, R.J. Next-Generation Biology: Sequencing and Data Analysis Approaches for Non-Model Organisms. *Mar. Genom.* **2016**, *30*, 3–13. [CrossRef] [PubMed]
15. O'Driscoll, A.; Daugelaite, J.; Sleator, R.D. 'Big Data', Hadoop and Cloud Computing in Genomics. *J. Biomed. Inform.* **2013**, *46*, 774–781. [CrossRef]

16. Maldonado, E. Bioinformatics Applications for Optimizing Downstream Analyses in Evolutionary Genomics. Ph.D. Thesis, Faculty of Sciences, University of Porto, Porto, Portugal, 2020. Available online: https://hdl.handle.net/10216/127527 (accessed on 2 June 2024).
17. Bao, R.; Huang, L.; Andrade, J.; Tan, W.; Kibbe, W.A.; Jiang, H.; Feng, G. Review of Current Methods, Applications, and Data Management for the Bioinformatics Analysis of Whole Exome Sequencing. *Cancer Inform.* **2014**, *13*, CIN.S13779. [CrossRef]
18. Boles, N.C.; Stone, T.; Bergeron, C.; Kiehl, T.R. Big Data Access and Infrastructure for Modern Biology: Case Studies in Data Repository Utility. *Ann. N. Y. Acad. Sci.* **2017**, *1387*, 112–123. [CrossRef]
19. Fisch, K.M.; Meissner, T.; Gioia, L.; Ducom, J.C.; Carland, T.M.; Loguercio, S.; Su, A.I. Omics Pipe: A Community-Based Framework for Reproducible Multi-Omics Data Analysis. *Bioinformatics* **2015**, *31*, 1724–1728. [CrossRef]
20. Hasin, Y.; Seldin, M.; Lusis, A. Multi-Omics Approaches to Disease. *Genome Biol.* **2017**, *18*, 83. [CrossRef]
21. Chen, R.; Mias, G.I.; Li-Pook-Than, J.; Jiang, L.; Lam, H.Y.; Chen, R.; Miriami, E.; Karczewski, K.J.; Hariharan, M.; Dewey, F.E.; et al. Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes. *Cell* **2012**, *148*, 1293–1307. [CrossRef]
22. Hawkins, R.D.; Hon, G.C.; Ren, B. Next-Generation Genomics: An Integrative Approach. *Nat. Rev. Genet.* **2010**, *11*, 476–486. [CrossRef]
23. Shi, L.; Wang, Z. Computational Strategies for Scalable Genomics Analysis. *Genes* **2019**, *10*, 1017. [CrossRef] [PubMed]
24. Stephens, Z.D.; Lee, S.Y.; Faghri, F.; Campbell, R.H.; Zhai, C.; Efron, M.J.; Iyer, R.; Schatz, M.C.; Sinha, S.; Robinson, G.E. Big Data: Astronomical or Genomical? *PLoS Biol.* **2015**, *13*, e1002195. [CrossRef] [PubMed]
25. Intel CPU vs. GPU: What's the Difference? Available online: https://www.intel.com/content/www/us/en/products/docs/processors/cpu-vs-gpu.html (accessed on 4 October 2024).
26. Phillips, J.C.; Hardy, D.J.; Maia, J.D.C.; Stone, J.E.; Ribeiro, J.V.; Bernardi, R.C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; et al. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130. [CrossRef] [PubMed]
27. Ahmad, T.; Ahmed, N.; Al-Ars, Z.; Hofstee, H.P. Optimizing Performance of GATK Workflows Using Apache Arrow In-Memory Data Framework. *BMC Genom.* **2020**, *21*, 683. [CrossRef]
28. Maldonado, E. MMAP: Multi-Core Method for Analysis Pipelines. Available online: https://mymmap.sourceforge.io/ (accessed on 18 June 2024).
29. Maldonado, E.; Antunes, A. LMAP_S: Lightweight Multigene Alignment and Phylogeny eStimation. *BMC Bioinform.* **2019**, *20*, 739. [CrossRef]
30. Maldonado, E.; Almeida, D.; Escalona, T.; Khan, I.; Vasconcelos, V.; Antunes, A. LMAP: Lightweight Multigene Analyses in PAML. *BMC Bioinform.* **2016**, *17*, 354. [CrossRef]
31. Ghedira, K.; Khamessi, O.; Hkimi, C.; Kamoun, S.; Dhamer, N.; Daassi, K.; Ben Salah, W.; Othman, H.; Belhadj, W.; Ghorbal, Y. Design and Implementation of a Scalable High-Performance Computing (HPC) Cluster for Omics Data Analysis: Achievements, Challenges and Recommendations in LMICs. *GigaScience* **2024**, *13*, giae060. [CrossRef]
32. Jambulingam, D.; Rathinakannan, V.S.; Heron, S.; Schleutker, J.; Fey, V. Kuura—An Automated Workflow for Analyzing WES and WGS Data. *PLoS ONE* **2024**, *19*, e0296785. [CrossRef]
33. Ahmad, T.; Al Ars, Z.; Hofstee, H.P. VC@Scale: Scalable and High-Performance Variant Calling on Cluster Environments. *GigaScience* **2021**, *10*, giab057. [CrossRef]
34. Leung, Y.Y.; Valladares, O.; Chou, Y.-F.; Lin, H.-J.; Kuzma, A.B.; Cantwell, L.; Qu, L.; Gangadharan, P.; Salerno, W.J.; Schellenberg, G.D.; et al. VCPA: Genomic Variant Calling Pipeline and Data Management Tool for Alzheimer's Disease Sequencing Project. *Bioinformatics* **2019**, *35*, 1768–1770. [CrossRef]
35. Chiara, M.; Gioiosa, S.; Chillemi, G.; D'Antonio, M.; Flati, T.; Picardi, E.; Zambelli, F.; Horner, D.S.; Pesole, G.; Castrignanò, T. CoVaCS: A Consensus Variant Calling System. *BMC Genom.* **2018**, *19*, 120. [CrossRef] [PubMed]
36. Reid, J.G.; Carroll, A.; Veeraraghavan, N.; Dahdouli, M.; Sundquist, A.; English, A.; Bainbridge, M.; White, S.; Salerno, W.; Buhay, C.; et al. Launching Genomics into the Cloud: Deployment of Mercury, a next Generation Sequence Analysis Pipeline. *BMC Bioinform.* **2014**, *15*, 30. [CrossRef] [PubMed]
37. Almeida, D.; Maldonado, E.; Khan, I.; Silva, L.; Gilbert, M.T.; Zhang, G.; Jarvis, E.D.; O'Brien, S.J.; Johnson, W.E.; Antunes, A. Whole Genome Identification, Phylogeny and Evolution of the Cytochrome P450 Family 2 (CYP2) Sub-Families in Birds. *Genome Biol. Evol.* **2016**, *8*, 1115–1131. [CrossRef] [PubMed]
38. Khan, I.; Yang, Z.; Maldonado, E.; Li, C.; Zhang, G.; Gilbert, M.T.; Jarvis, E.D.; O'Brien, S.J.; Johnson, W.E.; Antunes, A. Olfactory Receptor Subgenomes Linked with Broad Ecological Adaptations in Sauropsida. *Mol. Biol. Evol.* **2015**, *32*, 2832–2843. [CrossRef]
39. Alvelos, M.I.; Gonçalves, C.I.; Coutinho, E.; Almeida, J.T.; Bastos, M.; Sampaio, M.L.; Melo, M.; Martins, S.; Dinis, I.; Mirante, A.; et al. Maturity-Onset Diabetes of the Young (MODY) in Portugal: Novel GCK, HNFA1 and HNFA4 Mutations. *J. Clin. Med.* **2020**, *9*, 288. [CrossRef]
40. Lemos, M.C.; Thakker, R.V. Hypoparathyroidism, Deafness, and Renal Dysplasia Syndrome: 20 Years after the Identification of the First GATA3 Mutations. *Hum. Mutat.* **2020**, *41*, 1341–1350. [CrossRef]

41. Maldonado, E.; Khan, I. Omics Biology in Diagnosis of Diseases: Towards Empowering Genomic Medicine from an Evolutionary Perspective. *Life* **2024**, *14*, 1637. [CrossRef]
42. Maldonado, E.; Khan, I. (Eds.) Multi-Omics for Diagnosing Diseases: Bioinformatics Approaches and Integrative Data Analyses. *Computation* **2025**, *in press*. Available online: https://www.mdpi.com/journal/computation/special_issues/D97DPHGA83 (accessed on 11 March 2025).
43. Maldonado, E.; Khan, I. (Eds.) Multi-Omics for Diagnosing Diseases: Bioinformatics Approaches and Integrative Data Analyses. *Life* **2025**, *in press*. Available online: https://www.mdpi.com/journal/life/special_issues/15JSWLKS45 (accessed on 11 March 2025).
44. Fernandes, M.Z.; Caetano, C.F.; Gaspar, C.; Oliveira, A.S.; Palmeira-de-Oliveira, R.; Martinez-de-Oliveira, J.; Rolo, J.; Palmeira-de-Oliveira, A. Uncovering the Yeast Diversity in the Female Genital Tract: An Exploration of Spatial Distribution and Antifungal Resistance. *Pathogens* **2023**, *12*, 595. [CrossRef]
45. Caetano, C.F.; Gaspar, C.; Oliveira, A.S.; Palmeira-de-Oliveira, R.; Rodrigues, L.; Gonçalves, T.; Martinez-de-Oliveira, J.; Palmeira-de-Oliveira, A.; Rolo, J. Study of Ecological Relationship of Yeast Species with Candida Albicans in the Context of Vulvovaginal Infections. *Microorganisms* **2023**, *11*, 2398. [CrossRef] [PubMed]
46. Bonifácio, M.; Mateus, C.; Alves, A.R.; Maldonado, E.; Duarte, A.P.; Domingues, F.; Oleastro, M.; Ferreira, S. Natural Transformation as a Mechanism of Horizontal Gene Transfer in Aliarcobacter Butzleri. *Pathogens* **2021**, *10*, 909. [CrossRef] [PubMed]
47. University of Beira Interior C4-UBI—Cloud Computing Competence Centre, University of Beira Interior. Available online: https://c4.ubi.pt/ (accessed on 13 June 2024).
48. Connor, T.R.; Loman, N.J.; Thompson, S.; Smith, A.; Southgate, J.; Poplawski, R.; Bull, M.J.; Richardson, E.; Ismail, M.; Thompson, S.E.; et al. CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): An Online Resource for the Medical Microbiology Community. *Microb. Genom.* **2016**, *2*, e000086. [CrossRef] [PubMed]
49. Bagheri, H.; Muppirala, U.; Masonbrink, R.E.; Severin, A.J.; Rajan, H. Shared Data Science Infrastructure for Genomics Data. *BMC Bioinform.* **2019**, *20*, 436. [CrossRef]
50. Simonyan, V.; Chumakov, K.; Dingerdissen, H.; Faison, W.; Goldweber, S.; Golikov, A.; Gulzar, N.; Karagiannis, K.; Vinh Nguyen Lam, P.; Maudru, T.; et al. High-Performance Integrated Virtual Environment (HIVE): A Robust Infrastructure for next-Generation Sequence Data Analysis. *Database* **2016**, *2016*, baw022. [CrossRef]
51. Blitz, R.; Storck, M.; Baune, B.T.; Dugas, M.; Opel, N. Design and Implementation of an Informatics Infrastructure for Standardized Data Acquisition, Transfer, Storage, and Export in Psychiatric Clinical Routine: Feasibility Study. *JMIR Ment. Health* **2021**, *8*, e26681. [CrossRef]
52. Bakken, S. An Informatics Infrastructure Is Essential for Evidence-Based Practice. *J. Am. Med. Inform. Assoc.* **2001**, *8*, 199–201. [CrossRef]
53. Forghani, A.; Sadjadi, S.J.; Moghadam, B.F. A Supplier Selection Model in Pharmaceutical Supply Chain Using PCA, Z-TOPSIS and MILP: A Case Study. *PLoS ONE* **2018**, *13*, e0201604. [CrossRef]
54. Zolghadri, M.; Eckert, C.; Zouggar, S.; Girard, P. Power-Based Supplier Selection in Product Development Projects. *Comput. Ind.* **2011**, *62*, 487–500. [CrossRef]
55. Schrader, D. The Fundamentals of Network Access Management. 2024. Available online: https://blog.netwrix.com/ (accessed on 10 March 2024).
56. Gubin, A.V. ZFS RAIDZ vs. Traditional RAID. Available online: https://www.klennet.com/notes/2019-07-04-raid5-vs-raidz.aspx (accessed on 7 August 2023).
57. Patterson, D.A.; Gibson, G.; Katz, R.H. A Case for Redundant Arrays of Inexpensive Disks (RAID). In Proceedings of the 1988 ACM SIGMOD International Conference on Management of Data, Chicago, IL, USA, 1–3 June 1988; Association for Computing Machinery: New York, NY, USA, 1988; pp. 109–116.
58. iXsystems TrueNAS SCALE—Linux-Based Open Source Storage Infrastructure. Available online: https://www.truenas.com/truenas-scale/ (accessed on 13 June 2024).
59. Salter, J. ZFS Fans, Rejoice—RAIDz Expansion Will Be a Thing Very Soon. Available online: https://arstechnica.com/gadgets/2021/06/raidz-expansion-code-lands-in-openzfs-master/ (accessed on 7 August 2023).
60. Calomel.org ZFS Raidz Performance, Capacity and Integrity Comparison @ Calomel.Org. Available online: https://calomel.org/zfs_raid_speed_capacity.html (accessed on 7 August 2023).
61. Qiao, Z.; Fu, S.; Chen, H.-B.; Settlemyer, B. Building Reliable High-Performance Storage Systems: An Empirical and Analytical Study. In Proceedings of the 2019 IEEE International Conference on Cluster Computing (CLUSTER), Albuquerque, Mexico, 23–26 September 2019; pp. 1–10.
62. Qiao, Z.; Liang, S.; Fu, S.; Chen, H.-B.; Settlemyer, B. Characterizing and Modeling Reliability of Declustered RAID for HPC Storage Systems. In Proceedings of the 2019 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks—Industry Track, Portland, OR, USA, 24–27 June 2019; pp. 17–20.

63. OpenZFS dRAID—OpenZFS Documentation. Available online: https://openzfs.github.io/openzfs-docs/Basic%20Concepts/dRAID%20Howto.html (accessed on 7 August 2023).
64. Holland, M.; Gibson, G.A. Parity Declustering for Continuous Operation in Redundant Disk Arrays. *ACM SIGPLAN Not.* **1992**, *27*, 23–35. [CrossRef]
65. iXsystems ZFS Primer. Available online: https://www.truenas.com/docs/references/zfsprimer/ (accessed on 14 June 2024).
66. Infostor. LinuxCon: OpenZFS Moves Open Source Storage Forward—Infostor.com®. Infostor.com. 2013. Available online: https://www.infostor.com/storage-management/linuxcon-openzfs-moves-open-source-storage-forward.html (accessed on 13 June 2024).
67. iXsystems SCALE Hardware Guide. Available online: https://www.truenas.com/docs/scale/gettingstarted/scalehardwareguide/ (accessed on 7 August 2023).
68. Harris, R. Why RAID 5 Stops Working in 2009. Available online: https://www.zdnet.com/article/why-raid-5-stops-working-in-2009/ (accessed on 7 August 2023).
69. Harris, R. Why RAID 6 Stops Working in 2019. Available online: https://www.zdnet.com/article/why-raid-6-stops-working-in-2019/ (accessed on 7 August 2023).
70. iXsystems TrueNAS Community. Available online: https://www.truenas.com/community/ (accessed on 7 August 2023).
71. iXsystems ZFS dRAID Primer. Available online: https://www.truenas.com/docs/references/draidprimer/ (accessed on 13 June 2024).
72. IDEMA The Advent of Advanced Format | IDEMA. Available online: https://idema.org/?page_id=2369 (accessed on 7 August 2023).
73. Fitzpatrick, M.E. *4K Sector Disk Drives: Transitioning to the Future with Advanced Format Technologies*; Toshiba America Information Systems, Inc.: Irvine, CA, USA, 2011; p. 12.
74. Seagate Transition to Advanced Format 4K Sector Hard Drives | Seagate US. Available online: https://www.seagate.com/blog/advanced-format-4k-sector-hard-drives-master-ti/ (accessed on 7 August 2023).
75. Hong, J.; Lee, D.; Hwang, A.; Kim, T.; Ryu, H.-Y.; Choi, J. Rare Disease Genomics and Precision Medicine. *Genom. Inform.* **2024**, *22*, 28. [CrossRef] [PubMed]
76. Zandesh, Z.; Ghazisaeedi, M.; Devarakonda, M.V.; Haghighi, M.S. Legal Framework for Health Cloud: A Systematic Review. *Int. J. Med. Inf.* **2019**, *132*, 103953. [CrossRef] [PubMed]
77. Hekel, R.; Budis, J.; Kucharik, M.; Radvanszky, J.; Pös, Z.; Szemes, T. Privacy-Preserving Storage of Sequenced Genomic Data. *BMC Genom.* **2021**, *22*, 712. [CrossRef] [PubMed]
78. Cabading, Z. SAS vs SATA: What's the Difference? | HP® Tech Takes. Available online: https://www.hp.com/us-en/shop/tech-takes/sas-vs-sata (accessed on 7 August 2023).
79. Cook, J.D. Hard Disk Array Failure Probabilities. Available online: https://www.johndcook.com/blog/2009/01/05/rai-failure-probabilities/ (accessed on 7 August 2023).
80. Wood, D.E.; Lu, J.; Langmead, B. Improved Metagenomic Analysis with Kraken 2. *Genome Biol.* **2019**, *20*, 257. [CrossRef]
81. Wood, D.E.; Salzberg, S.L. Kraken: Ultrafast Metagenomic Sequence Classification Using Exact Alignments. *Genome Biol.* **2014**, *15*, R46. [CrossRef]
82. Gouy, M.; Guindon, S.; Gascuel, O. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol. Biol. Evol.* **2010**, *27*, 221–224. [CrossRef]
83. Gouy, M.; Tannier, E.; Comte, N.; Parsons, D.P. Seaview Version 5: A Multiplatform Software for Multiple Sequence Alignment, Molecular Phylogenetic Analyses, and Tree Reconciliation. In *Multiple Sequence Alignment: Methods and Protocols*; Katoh, K., Ed.; Springer: New York, NY, USA, 2021; pp. 241–260, ISBN 978-1-07-161036-7.
84. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative Genomics Viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [CrossRef]
85. Jaganathan, K.; Kyriazopoulou Panagiotopoulou, S.; McRae, J.F.; Darbandi, S.F.; Knowles, D.; Li, Y.I.; Kosmicki, J.A.; Arbelaez, J.; Cui, W.; Schwartz, G.B.; et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **2019**, *176*, 535–548.e24. [CrossRef]
86. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [CrossRef]
87. Guindon, S.; Dufayard, J.F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [CrossRef]
88. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and Applications. *BMC Bioinform.* **2009**, *10*, 421. [CrossRef] [PubMed]
89. Grüning, B.; Dale, R.; Sjödin, A.; Chapman, B.A.; Rowe, J.; Tomkins-Tinch, C.H.; Valieris, R.; Köster, J. Bioconda: Sustainable and Comprehensive Software Distribution for the Life Sciences. *Nat. Methods* **2018**, *15*, 475–476. [CrossRef] [PubMed]

90. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]
91. Barnett, D.W.; Garrison, E.K.; Quinlan, A.R.; Strömberg, M.P.; Marth, G.T. BamTools: A C++ API and Toolkit for Analyzing and Managing BAM Files. *Bioinformatics* **2011**, *27*, 1691–1692. [CrossRef]
92. Katz, K.; Shutov, O.; Lapoint, R.; Kimelman, M.; Brister, J.R.; O'Sullivan, C. The Sequence Read Archive: A Decade More of Explosive Growth. *Nucleic Acids Res.* **2022**, *50*, D387–D390. [CrossRef]
93. Maurya, A.; Szymanski, M.; Karlowski, W.M. ARA: A Flexible Pipeline for Automated Exploration of NCBI SRA Datasets. *GigaScience* **2023**, *12*, giad067. [CrossRef]
94. Andrews, S. FastQC A Quality Control Tool for High Throughput Sequence Data. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 13 February 2025).
95. Magoč, T.; Salzberg, S.L. FLASH: Fast Length Adjustment of Short Reads to Improve Genome Assemblies. *Bioinformatics* **2011**, *27*, 2957–2963. [CrossRef]
96. Krueger, F. Trim Galore!: A Wrapper around Cutadapt and FastQC to Consistently Apply Adapter and Quality Trimming to FastQ Files, with Extra Functionality for RRBS Data. Available online: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (accessed on 13 February 2025).
97. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef]
98. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef]
99. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [CrossRef]
100. Kolmogorov, M.; Raney, B.; Paten, B.; Pham, S. Ragout—A Reference-Assisted Assembly Tool for Bacterial Genomes. *Bioinformatics* **2014**, *30*, i302–i309. [CrossRef] [PubMed]
101. Dierckxsens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: De Novo Assembly of Organelle Genomes from Whole Genome Data. *Nucleic Acids Res.* **2017**, *45*, e18. [CrossRef] [PubMed]
102. Boras, M.; Balen, J.; Vdovjak, K. Performance Evaluation of Linux Operating Systems. In Proceedings of the 2020 International Conference on Smart Systems and Technologies (SST), Osijek, Croatia, 14–16 October 2020; pp. 115–120.
103. Balen, J.; Vajak, D.; Salah, K. Comparative Performance Evaluation of Popular Virtual Private Servers. *J. Internet Technol.* **2020**, *21*, 343–356.
104. CPUID POWERMAX | Softwares | CPUID. Available online: https://www.cpuid.com/softwares/powermax.html (accessed on 27 August 2024).
105. CPUID CPU-Z | Softwares | CPUID. Available online: https://www.cpuid.com/softwares/cpu-z.html (accessed on 27 August 2024).
106. Miyazaki, N. CrystalDiskMark. Available online: https://sourceforge.net/projects/crystaldiskmark/ (accessed on 27 August 2024).
107. Phoronix Media Phoronix Test Suite—Linux Testing & Benchmarking Platform, Automated Testing, Open-Source Benchmarking. Available online: https://www.phoronix-test-suite.com/ (accessed on 27 August 2024).
108. Primate Labs Inc. Geekbench 6—Cross-Platform Benchmark. Available online: https://www.geekbench.com/ (accessed on 27 August 2024).
109. Ultimate Systems HARDiNFO Benchmark—The Ultimate Hardware Performance Benchmarks. Available online: https://www.hardinfo-benchmark.com/ (accessed on 27 August 2024).
110. Dugan, J.; Elliott, S.; Mah, B.A.; Poskanzer, J.; Prabhu, K.; Ashley, M.; Brown, A.; Jaißle, A.; Sahani, S.; Simpson, B.; et al. iPerf—The TCP, UDP and SCTP Network Bandwidth Measurement Tool. Available online: https://iperf.fr/ (accessed on 27 August 2024).
111. Lucas, K. Kdlucas/Byte-Unixbench—A Unix Benchmark Suite. Available online: https://github.com/kdlucas/byte-unixbench (accessed on 27 August 2024).