

Genetic Data Analysis and Business Process Management Platform for Personalized Nutrition Service

Jitao Yang(✉)

School of Information Science, Beijing Language and Culture University, Beijing 100083, China
yangjitao@blcu.edu.cn

Abstract. Vitamins and minerals are essential micro-nutrients required by our bodies to optimize health and maintain well-being. Currently, our nutrient recommendations are one-size-fits-all or one solution for a large group of persons, however, due to the difference of each individual's genetic background, lifestyle, and health condition, individual's nutrition requirement is different from each other, therefore, we should compute and provide personalized nutrition supplementation solution for each person. At the beginning of this paper, we describe the importance of personalized nutrition for keeping healthy; then we give a description of our personalized nutrition intelligent service platform by analyzing genetic data, lifestyle data and physical examination data together to generate genetic interpretation report and personalized nutrition report, and further place order to produce customized nutrition packs for each customer. To promote management efficiency and reduce errors in business processes, we developed multiple business process management systems such as laboratory information management system (LIMS), bioinformatic pipelines, genetic interpretation system, customer relationship management (CRM) system and etc. All the systems were integrated together to have the ability of processing tens of thousands of samples in parallel.

Keywords: Genetic testing · Nutrigenetics · Personalized nutrition platform · Business process management

1 Introduction

The traditional nutrient recommendations: Dietary Reference Intakes (DRIs) [1] are estimated average daily nutrient intake levels to meet the requirements of a group of population.

However, the difference of genetic background will affect the digestion, absorption, transport, metabolism and storage of food and nutrients, and the steady-state regulation of internal environment to varying degrees, reflecting the differences of different individual's response to nutrition and susceptibility to disease.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
O. Gervasi et al. (Eds.): ICCSA 2022 Workshops, LNCS 13381, pp. 550–559, 2022.
https://doi.org/10.1007/978-3-031-10548-7_40

Nutrigenetics can study how individual's genetic composition affects diet response and susceptibility to diet related diseases through retrospective analysis of the interaction between individual's genetic variations and diet as well as diseases.

As a new interdisciplinary subject born from the benefit of the human genome project, nutrigenetics has produced many scientific research achievements, such as individual's dietary behaviors, vitamin and other nutritional needs, as well as metabolic factor regulation, are all significantly different due to the composition of individual's genetic background. Genetic defects may cause the body to be unable to obtain nutrition from food, then health will be damaged, therefore, it is necessary to evaluate nutrient requirements according to genetic composition, so as to make more accurate nutritional supplementation.

Vitamin D, is a fat soluble vitamin, and it's a necessary nutrient to maintain the life of human. Its main forms are vitamins D_3 and D_2, which form active vitamin D through hydroxylation in liver and kidney, and then exert its biological activity. Vitamin D is hydroxylated to 25-hydroxy vitamin D at position 25 and is further hydroxylated to 1,25-dihydroxy vitamin D [1,25 (OH) 2D] by 1α-Hydroxylase (encoded by the gene CYP27B1) in the kidney; vitamin D binding protein (DBP) binds to vitamin D for transport, 1,25 (OH) 2D is released into the blood circulation of the intestine, and then through the mechanism mediated by vitamin D receptor (encoded by the gene VDR), passively reabsorbed in the kidney [2].

Nutrigenetics research shows that the genetic polymorphisms of multiple genes involved in vitamin D metabolic pathway affect the activation, transport and absorption of vitamin D. For example, the serum concentration of 1,25(OH)2D3 in hepatitis C virus (HCV) patients with CYP27B1-1260 AA genotype is higher than that in HCV patients with CYP27B1-1260 AC or CC genotype, and the sustained virological response rate (SVR) of the HCV patients with CYP27B1-1260_AA homozygous genotype was significantly increased [3].

The GC gene's mutation can encode different GC protein mutants, the Gc2-2 protein mutant encoded by 436KK mutation has the lowest concentration of vitamin D binding protein in individual's plasma [4].

As a ligand, active vitamin D binds to vitamin receptor (VDR) and enters into the cell, and forms a complex with retinoid X receptor (RXR) in the nucleus. RXR-VDR binds to vitamin D response element (VDRE) and affects gene expression regulation as a transcription factor [5].

The Cdx2 gene's polymorphisms are located in the promoter region of VDR gene, allele A can improve the transcriptional activity of VDR gene. The FokI gene's restriction enzyme cutting site is located in the coding region of VDR gene, which has an impact on the activity of receptor. The BsmI, ApaI and TaqI genes' restriction enzyme cutting sites are located at the 3' end of VDR gene. The VDR gene's polymorphisms can limit the process of vitamin D entering cells [6].

With the development and application of functional genomics and next-generation sequencing technology, the time and cost of whole genome sequencing

(WGS) are greatly shortened. The establishment of genomics database based on population cohort research, the coding and interpretation of basic genome information by bioinformatics analysis technology, and the use of polygenic risk scoring (PRS) model, make the application of individual genome information possible, which can provide effective technical support for the accurate evaluation of individual's health status and the formulation of intervention strategies.

In recent years, bioinformatics has become more and more mature, and the application of single nucleotide polymorphism (SNP) based genetic risk score (GRS), PRS and other genetic risk assessment models have gradually expanded, more and more studies have focused on the field of nutrigenetics, such as the GRS based vitamin D insufficiency genetic risk assessment model [7], the risk allele frequency based vitamin D deficiency genetic risk assessment model [8], the intervention effect evaluation [9], and the evaluation model of vitamin E supplement response based on partial least squares (PLS) regression model [10].

2 Genetic Data Analysis

To analyze a customer's genetic data, we provide an non-invasive and self-administered DNA collection solution that, we send customer a DNA collection kit, which includes a saliva collecting tube, in the tube is the stabilization chemistry reagent used to keep the stabilization of saliva and DNA. After spiting saliva to the saliva collection tube, the customer should post the DNA collection kit to our molecular genetic testing laboratory. The laboratory will extract DNA from the saliva, and after a series of complex experimental processes, the DNA samples will be sequenced by different genome sequencing equipment such as iScan [11], Novaseq [12] and etc. The sequencer will generate millions of reads data, and the gene sequencing data will be stored in cloud storage disks. Multiple bioinformatic pipelines will analyze the reads data and give the genetic risk scores. Based on the genetic risk scores, the genetic interpretation system will generate genetic interpretation report using the gene-disease associations databases such as the databases of AutDB [13], DisGeNET [14,15], OMIM [16,17], and other published clinical scientific literature.

Figure 1 left is the user interface of precision nutrition genetic testing report, which lists the nutrition related genetic testing items, such as Vitamin A (VA), Calcium, Vitamin C (VC), Iron, Vitamin D (VD), Zinc, Vitamin E (VE), Selenium, Vitamin K (VK), Magnesium, Vitamin B_2 (VB_2), Phosphorus, Folic acid, Vitamin B_3 (VB_3), Lactose Tolerance, Vitamin B_{12} (VB_{12}), Alcohol Metabolism, Vitamin B_6 (VB_6), Caffeine Metabolism, Omega-3 Intake, Sweetness Sensitivity, Bitter Sensitivity, and etc. Under the genetic testing item name, the color bar represents the customer's nutrition requirement levels (*i.e.*, high, higher, medium, low). Click the name of each genetic testing item in the report, more detail information will be displayed to: 1) explain the genetic testing result, 2) give life style suggestions, 3) recommend personalized nutrition solutions, 4) list loci's genotypes, and 5) provide scientific evidences.

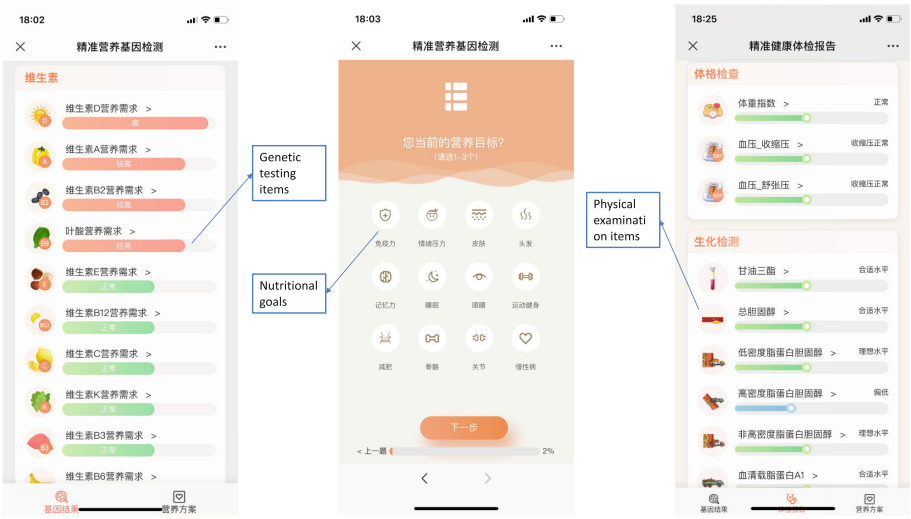

Fig. 1. The user interfaces of personalized nutrition service reports.

The genetic interpretation reports are embedded in WeChat app using the APIs [18] of WeChat. By embedding the genetic interpretation reports in WeChat, we can reach and communicate with customers conveniently, rather than persuade customers to download a new app or open a web page to access the genetic report. Since WeChat is the most popular chatting software in China, and more than 1 billion people in China use WeChat for communication everyday, therefore, customers can find and read the genetic interpretation reports easily, and we can also send notifications to customers through WeChat for notifying the status of samples.

3 Personalized Nutrition Service

To get customers' lifestyle data, we provide online food frequency questionnaires (FFQs) [19–21] to collect customers' diet, sports, sleep, and other health related data, including the questions related to health status/physical conditions, dietary habits, and medical histories. Figure 1 middle is the user interface of online food frequency questionnaires. The questions in the online FFQs are dynamic that, at the beginning, the questionnaire will list multiple nutritional goals, based on different nutritional goals selected by customers, the system will generate different questions, and the answer of the current question can decide the content of next question.

Additionally, we provide an interface for customers to upload their physical examination report to our platform, including the physical examination data of: heart rate, complete blood count, throat, urinalysis, ears, blood pressure, respiration rate, thyroid, nose, eyes, and etc. Figure 1 right is the user interface of

physical examination report. Since different physical examination centers have their different physical examination report formats in PDF, therefore, our platform has the function to adjust to different PDF formats to extract physical examination data.

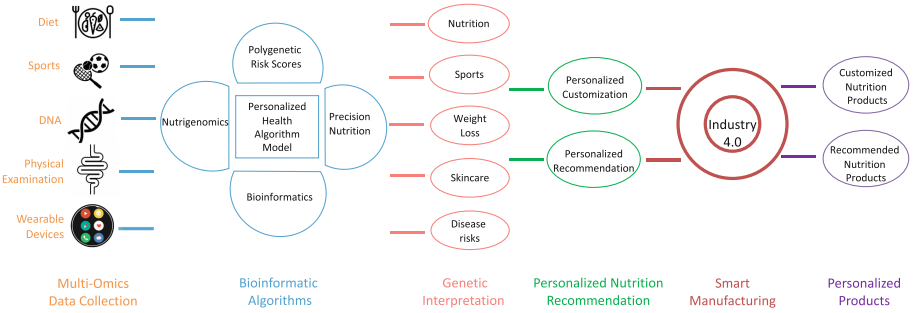

Fig. 2. Personalized nutrition intelligent service platform.

Figure 2 demonstrates the personalized nutrition intelligent service (PNIS) platform. Combing the genetic testing data, lifestyle data, physical examination data, PNIS platform will analyze the data together using bioinformatic algorithms, genetic interpretation algorithms, and nutrition requirement evaluation algorithms, to generate genetic interpretation reports, and further generate personalized nutrition report. Based on the personalized nutrition report, PNIS platform can order smart manufacturing machine in nutrition factory directly to produce unique customized nutrition product for each customer.

Genome sequencing has multiple platforms, different sequencing platform requires different laboratory experimental methods and process, and generate the sequence data in different formats. PNIS platform can start different bioinformatic pipelines to execute quality control and analyze the genomic sequencing data from different genome sequencing equipment.

PNIS platform can analyze tens of thousands of genetic testing samples in parallel, the bioinformatic analysis pipelines were implemented in cloud environment and scheduled by Kubernetes [22]. Each bioinformatic analysis software was en-capsuled in a docker engine, and multiple bioinformatic analysis software dockers can be chained together to form pipelines. When there are large amount of samples need to be analyzed, the platform will start more computing nodes elastically.

PNIS can generate multiple genetic interpretation reports, including precision nutrition, precision sports, weight loss, precision skincare, disease risks, mother care, men's health care, strengthen immune system, and anti-fatigue.

Based on the nutrigenetics, dietary guidelines, population nutrition big data, and customer's personal data, PNIS will compute and give a personalized

nutrition report for each person. The personalized nutrition report includes two solutions: personalized customization, and personalized recommendation.

The personalized customization solution will order the smart manufacturing nutrition factory to produce personalized and customized daily nutritional supplement packs for each customer. Figure 3 left is the user interface that lists the recommended daily nutritional supplements; Fig. 3 middle shows the nutrition packs that going to be produced by factory, every user's name will be printed on the nutrition pack and our system allows each customer to define his/her name on the pack, the name of the tablets and capsules inside the nutrition pack will also be printed on the pack so that to let customer know the ingredients of the nutrition pack.

The personalized nutrition packs will be produced and then posted to user's home or office directly from the nutrition factory. The personalized recommendation solution lists the suggested foods that can help customer to supplement the nutrients that customer's body required, as described in Fig. 3 right.

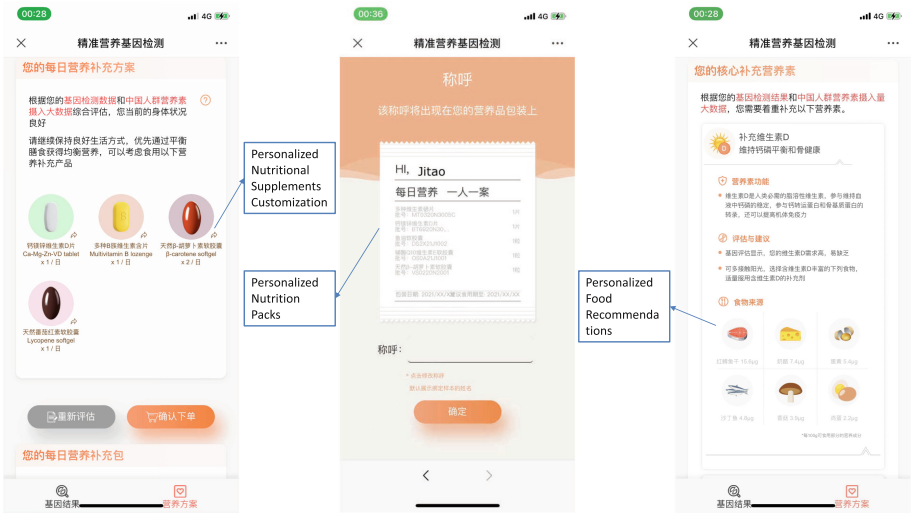

Fig. 3. The user interfaces of personalized customization and personalized recommendation reports.

4 Personalized Nutrition Service Business Process Management

To support the processing of large amount of samples in parallel and the collaboration with multiple business partners, we developed several systems including laboratory information management system (LIMS), bioinformatic pipelines scheduling system, genetic interpretation system, personalized nutrition report

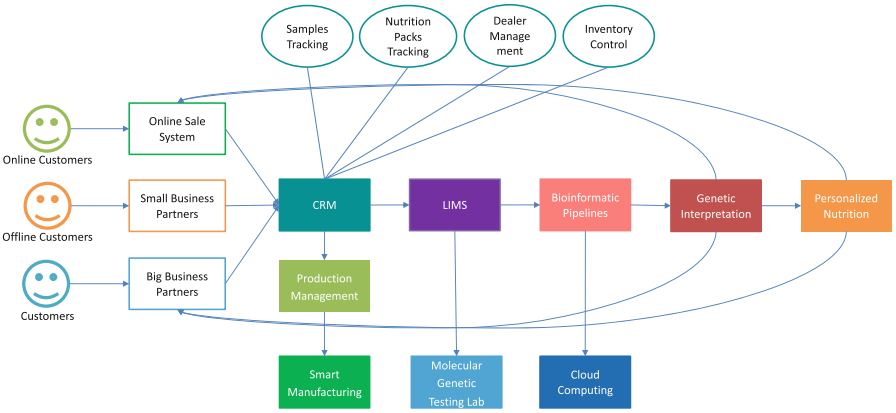

Fig. 4. Business process management systems for personalized nutrition service.

system, production management system, online sale system, and customer relation management system (CRM). Our business process is described in Fig. 4.

The online sale system has the functions of products display, placing order, shopping cart, multiple addresses management, order status tracking, logistics tracking, online payment, coupons, and return of goods.

The customer relationship management system provides both business to customer (B to C) service and business to business (B to B) service. The business to customer service connects with the online sale system, managing customer's personal information, order information, after sale service, as well as providing genetic testing sample tracking information (including the logistics information of sample mailed back to the molecular genetic testing laboratory, sample experiment progress information, and genetic interpretation report generation progress information) and personalized nutrition packs' production progress information (including manufacturing factory production progress information and logistics information of the personalized nutrition packs posted to customer's home or office). The business to business service can provide a customized user interface (using business partner's logo) for small business partners. The business to business service also provides APIs to communicate with big business partner's sale system to 1) provide sample tracking information, genetic interpretation reports, and personalized nutrition packs production information, as well as 2) exchange other information with big partner's system.

The LIMS manages all the sample experimental procedures such as sample reception, DNA extraction, library construction, polymerase chain reaction (PCR) [29] amplification, pooling, sequencing, sample tracking, and etc.

The bioinformatic pipelines scheduling system manages and schedules multiple bioinformatic pipelines to analyze large amount of genetic testing samples in parallel.

The genetic interpretation system interprets the analyzed genomic data, and generates different genetic interpretation reports such as precision nutrition

genetic interpretation report, weight loss genetic interpretation report, precision skincare genetic interpretation report, and so on based on customer's order.

The personalized nutrition report system can generate tailored nutrition report for each person combing customer's genetic data, lifestyle data, and physical examination data.

The production management system will receive personalized nutrition packs production order from CRM system and then place the order (including customer's name, tablet and capsule information, customer's address information) to smart manufacturing nutrition factory to produce daily personalized nutrition packs. After production, the daily personalized nutrition packs will be mailed to user's home or office directly.

As decribed in Fig. 4, the systems are integrated together and can exchange data conveniently between each other. Each system was developed using Spring Boot [23], MyBatis [24], BootStrap [26], Html5, CSS [27], and javascript, and implemented as micro services in Spring Cloud [28]. All the systems have backup and standby systems to stabilize the service for personalized nutrition solution.

5 Conclusions

Currently, most of the nutrition products are one-size-fits-all, and the traditional nutrient recommendations are estimated to meet the requirements of a large group of population.

However, studies found that genetic background will affect the digestion, absorption, transport, metabolism and storage of food and nutrients; additionally, individual's lifestyles and health conditions also have impact on individual's nutrition requirement. Therefore, different individuals should supplement different kinds and different amounts of foods and nutrients.

In this paper, based on nutrigenetics and nutriology, we developed a personalized nutrition intelligent service (PNIS) platform. PNIS collects individual's genetic data through genetic testing and bioinformatic analysis, harvests lifestyle data through food frequency questionnaires, and retrieves individual's physical examination data through individual's uploading of physical examination report. Through bioinformatic analysis and genetic interpretation, PNIS generates genetic interpretation reports for each person, and further combines lifestyle data and physical examination data to analyze and generate personalized nutrition reports. Based on personalized nutrition report, the smart manufacturing nutrition factory will produce daily customized nutrition packs for each consumer. To verify the effectiveness of our personalized nutrition solution, we have collaborated with a physical examination center to conduct a population cohort study, after personalized nutrition supplementation, the result showed that our personalized nutrition solution could effectively supplement the nutritional needs of consumers.

We also developed multiple information systems such as online sale system, CRM system, LIMS and so on to support the business process management of large amount of samples. The systems have been delivered online and can provide service for tens of thousands of customers in parallel in one day.

We will continue to optimize our personalized nutrition solution by analyzing more individual's personal data such as the sensor data of sports, sleep, and etc.

Acknowledgment. This research project is supported by Science Foundation of Beijing Language and Culture University (supported by "the Fundamental Research Funds for the Central Universities") (Approval number: 22YJ080008)

References

1. Dietary Reference Intakes (DRIs). https://www.nal.usda.gov/fnic/dietary-reference-intakes. Accessed 2 May 2022
2. Wu, X., Cheng, J., Yang, K.: Vitamin D-related gene polymorphisms, plasma 25-Hydroxy-Vitamin D, cigarette smoke and non-small cell lung cancer (NSCLC) risk. Int. J. Mol. Sci. **17**(10), 1597 (2016)
3. Lange, C.M., Bojunga, J., Ramos-Lopez, E., et al.: Vitamin D deficiency and a CYP27B1-1260 promoter polymorphism are associated with chronic hepatitis C and poor response to interferon-alfa based therapy. J. Hepatol. **54**(5), 887–893 (2011)
4. Malik, S., Fu, L., Juras, D.J., et al.: Common variants of the vitamin D binding protein gene and adverse health outcomes. Crit. Rev. Clin. Lab. Sci. **50**(1), 1–22 (2013)
5. Irimie, A.I., Braicu, C., Pasca, S., et al.: Role of key micronutrients from nutrigenetic and nutrigenomic perspectives in cancer prevention. Medicina (Kaunas) **55**(6), 283 (2019)
6. Wjst, M.: Variants in the vitamin D receptor gene and asthma. BMC Genet. **6**(2) (2005)
7. Wang, T.J., Zhang, F., Richards, J.B., et al.: Common genetic determinants of vitamin D insufficiency: a genome-wide association study. Lancet **376**(9736), 180–188 (2010)
8. Yoon, B.W., Shin, H.T., Seo, J.: Risk allele frequency analysis of single-nucleotide polymorphisms for Vitamin D concentrations in different ethnic group. Genes **12**(10), 1530 (2021)
9. Sallinen, R.J., Dethlefsen, O., Ruotsalainen, S., et al.: Genetic risk score for serum 25-Hydroxyvitamin D concentration helps to guide personalized Vitamin D supplementation in healthy Finnish adults. J. Nutr. **151**(2), 281–292 (2021)
10. Borel, P., Desmarchelier, C., Nowicki, M., Bott, R., Tourniaire, F.: Can genetic variability in α-tocopherol bioavailability explain the heterogeneous response to α-tocopherol supplements? Antioxid. Redox Signal. **22**(8), 669–678 (2015)
11. iScan System. https://www.illumina.com/systems/array-scanners/iscan.html. Accessed 2 May 2022
12. Novaseq. https://www.illumina.com/systems/sequencing-platforms/novaseq.html. Accessed 2 May 2022
13. AutDB: a Genetic Database for Autism Spectrum Disorders. http://www.mindspec.org/products/autdb/. Accessed 8 May 2022
14. DisGeNET - a database of gene-disease associations. https://www.disgenet.org/dbinfo. Accessed 8 May 2022
15. Pinero, J., Bravo, A., Queralt-Rosinach, N., Gutierrez-Sacristan, A., et al.: DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. **45**(D1), D833–D839 (2017)

16. OMIM-Online Mendelian Inheritance in Man. https://www.omim.org/. Accessed 8 May 2022
17. Amberger J. S., Hamosh A.: Searching online Mendelian inheritance in man (OMIM): A knowledgebase of human genes and genetic phenotypes. Curr. Protoc. Bioinform. **58**, 1.2.1–1.2.12 (2017)
18. WeChat open platform. https://open.weixin.qq.com/. Accessed 8 May 2022
19. Johns, R., Kusuma, J., Lie, A., Shiao, S.P.K.: Validation of macro- and micronutrients including methyl donors in social ethnic diets using food frequency questionnaire and nutrition data system for research (USDA computerized program). SDRP J. Food Sci. Technol. **3**(4), 417–430 (2018)
20. Affret, A., Fatouhi, D.E., Dow, C., Correia, E., Boutron-Ruault, M.C., Fagherazzi, G.: Relative validity and reproducibility of a new 44-Item diet and food frequency questionnaire among adults: online assessment. J. Med. Internet Res. **20**(7), e227 (2018)
21. Thompson, F.E., Subar A.F.: Chapter 1 - Dietary Assessment Methodology. Nutrition in the Prevention and Treatment of Disease, 4th edn. Academic Press, ISBN: 9780128029282 (2017)
22. Kubernetes. https://kubernetes.io/. Accessed 8 May 2022
23. Spring Boot. https://spring.io/projects/spring-boot Accessed 10 May 2022
24. MyBatis. https://mybatis.org/mybatis-3/. Accessed 10 May 2022
25. HTML Living Standard. https://html.spec.whatwg.org/. Accessed 10 May 2022
26. What is Bootstrap? https://www.w3schools.com/whatis/whatis_bootstrap.asp. Accessed 10 May 2022
27. Starting with HTML + CSS. https://www.w3.org/Style/Examples/011/firstcss.en.html. Accessed 10 May 2022
28. Spring Cloud. https://spring.io/projects/spring-cloud. Accessed 10 May 2022
29. Lorenz T.C.: Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. J. Vis. Exp. (63), e3998 (2012)