CROATIAN MEDICAL JOURNAL
CMJ
42(3):252-255,2001

FORENSIC SCIENCES

Automation and High Through-put for a DNA Database Laboratory: Development of a Laboratory Information Management System

Martin Steinlechner, Walther Parson

Institute of Legal Medicine, University of Innsbruck, Innsbruck, Austria

Automation and high through-put production of DNA profiles has become a necessity in every DNA database unit. In our laboratory we developed a Laboratory Information Management System (LIMS) controlled workflow architecture, which comprises a robotic DNA extraction- and pipetting-system and a capillary electrophoresis unit. This allows a through-put of 4,000 samples per person per year. Improved sample handling and data management, full sample- and batch-histories, and software-aided supervision of result data, with a consequent average turn-around time of 8 days, are the main features of our new system.

Key words: *Austria; automatic data processing; computers; database, factual; database management systems; DNA; forensic medicine; information system; information storage and retrieval; management information systems; software*

When the Austrian DNA Intelligence Database was launched by the Ministry of the Interior in October 1997, our laboratory was assigned the responsibility for DNA-typing of reference and casework samples from the entire Austria (1). The expected 12,000 reference samples per year required a drastic change in our laboratory set-up. Before implementing the automation, the laboratory strategy was mainly based on manual steps, slightly supported by a simple administrative software tool. The laboratory was equipped with a single gel electrophoresis unit. This served our needs, since we usually analyzed approximately 600 casework samples per year. Our new objective was to analyze 10,000 to 20,000 samples per year with maximum quality and efficiency, at minimum cost, and within minimum response time (2-4).

Firstly, the current processing was evaluated to determine steps that would benefit most from automation. The next phase was to work out a new strategy to establish an automated process with high sample through-put characteristics, a strategy with focus on integration of a robotic system and the use of capillary electrophoresis units.

There were two aims. The first was to develop laboratory methods to facilitate a lean, fast, and safe laboratory process using the instruments mentioned above (5). The second aim, which is discussed in this paper, was to develop a new software-based laboratory information management system to avoid administrative overhead and redundancy, and to integrate different instruments into an automated workflow architecture. This new system was to improve sample handling and data management (4).

Methods

Overview of the Laboratory Information Management System

The integrated system covers laboratory procedures like those in the original manual approach (extraction, amplification, separation by electrophoresis, and data analysis) and procedures for assembling kits for sample collection, sample- and batch-handling, checking result data, printing documentation, and transmitting DNA-profiles into the national DNA database. The heart of the integrated system is GerichtsMedizin Innsbruck Laboratory Information Management System (GMI-LIMS) software package (Fig. 1).

Basics of the Laboratory Information Management System

The hardware of the integrated system is quite heterogeneous. On the one hand, there is Apple Macintosh hardware for the control of the capillary electrophoresis unit and for DNA data analysis. On the other hand, there is PC hardware used for laboratory administration. Both hardware paltforms are networked via a Novell Netware 5.0-based server system, which makes available the basic communication features necessary for the software-based LIMS. A backup server system is also installed.

The software basis is GMI-LIMS, a Microsoft Access 2.0-based supervisory software package, which allows execution of and communication with proprietary software packages, such as Easylabel 1.9c for bar code printing (Tharo Systems, Brunswick, OH, USA), WinRUFAS 1.12 for the control of Plato 3000 robotic system (Rosys/Anthos AG, Hombrechtikon, Switzerland), ABI Prism 310 Data Collection for the control of CE 310 Genetic Analyzer (PE Biosystems, Foster City, CA, USA), ABI Prism Genescan 2.1 for DNA fragment analysis (PE Biosystems) and ABI Prism Genotyper 2.0 for allele designation (PE Biosystems). A complex referential data model has been designed to integrate all the different types of data for complete sample- and batch-history, and to achieve a software-controlled setting for the automated instruments. All printed types of documentation are generated by GMI-LIMS. We have implemented a sophisticated user hierarchy, with authorized access to the different levels of functionality and

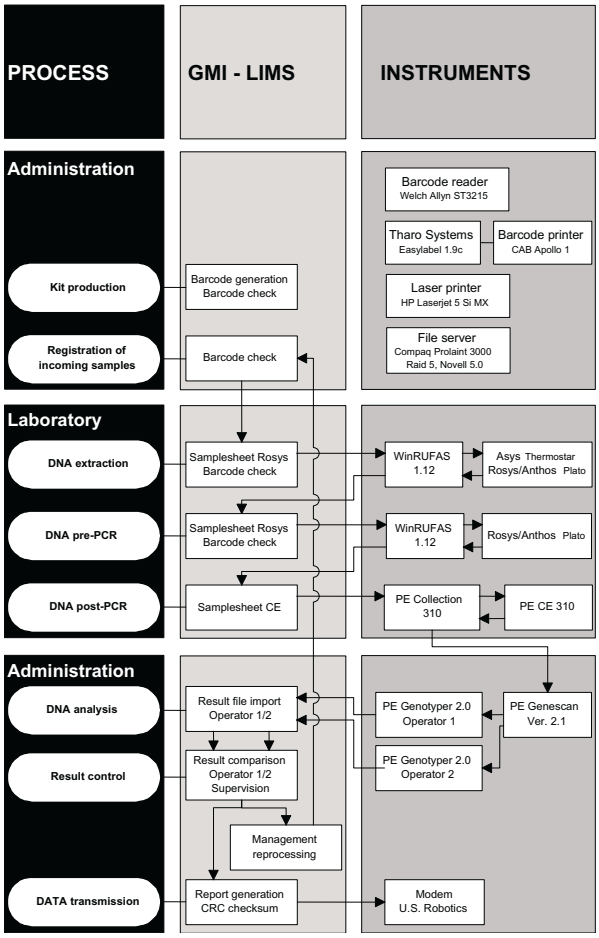

Figure 1. Overview of integrated GerichtsMedizin Innsbruck Laboratory Information Management System (GMI-LIMS).

interactions between logged users for quality assurance purposes. The user is guided through the workflow by GMI-LIMS, which prompts the user for all necessary actions and points out the errors.

Sample identification is supported by the bar code technology, namely, the "interleaved 2 of 5" symbols. "Herma PE weiß matt" (Herma, Stuttgart, Germany) with special glue (Herma) was chosen as carrier material for bar code labels. We use thermo-transfer bar code print technology with a CAB Apollo 1 printer and artificial resin transfer film (Herma). The bar codes are read by an industrial standard contact scanner (ST3215, Welch Allyn, Skaneateles Falls, NY, USA). The unique identifier is an 8-digit number; the first digit indicates the type of series (e.g., 9 means reference sample), the last digit is a check digit calculated by an algorithm using module ten.

Kit Production

Before kits for sample collection are assembled, GMI-LIMS generates a new bar code series. Four bar codes are printed for each kit and then put on the envelope, the ID-form, and two reaction tubes – all components of the kit. During assembly of each kit, bar coded components are scanned with the bar code reader. GMI-LIMS checks that the every kit component is labeled with the same unique bar code. GMI-LIMS also logs to which police station a kit has to be distributed.

Sample- and Batch-identification and Batch-generation

When the police returns two bar code-labeled reaction tubes each containing one buccal scrape to the laboratory, the tubes are scanned with the bar code reader. GMI-LIMS controls whether the unique identifier and the check digit of the bar code are valid, and then defines an archive position for one of the two tubes that is used for duplicate analysis in case of a match. The other tube is put on a plate for samples to be processed. After that, a batch can be set up.

The usual batch size is 48, comprising 43 reference samples plus five controls. For each batch GMI-LIMS generates a unique batch identifier, which is printed as a bar code. A specific position is defined for each sample in the batch. All plates necessary for the following procedures are then labeled with the batch identifier. After all samples of one batch are arranged, the batch identifiers on the plates and the bar codes on the sample tubes are scanned with the bar code reader. GMI-LIMS checks the sequence of all samples in the batch and whether they are correctly arranged.

The system allows processing of two batches (equal to 96-well format) in one run.

DNA Extraction, Polymerase Chain Reaction Set-up

DNA extraction (5) and polymerase chain reaction (PCR) (5) set-up are performed on a custom built robotic platform based on a 4-channel robotic microplate processor (Plato 3000, Rosys/Anthos) in combination with an extraction unit (Thermostar, Asys Hitech GmbH, Eugendorf, Austria). Both units are controlled by WinRUFAS software, which is executed by GMI-LIMS and provides sample-sheet and data. The batch identifiers and the sample bar codes are scanned with the bar code reader after plates and samples have been arranged on the platform. GMI-LIMS checks their validity and then executes WinRUFAS for DNA extraction and PCR set-up procedures. Log-data are recorded by WinRUFAS during the whole process. After that, the PCR plate is ready for PCR analysis.

Polymerase Chain Reaction

The PCR plate is placed on a thermal cycler (PE 9600) manually (5).

Capillary Electrophoresis

Amplification products, formamide, and internal lane standard are pipetted in a 96-well format with a 96-tip pipettor (Multispense, Asys Hitech GmbH). After denaturation, the plate is manually transferred to the capillary electrophoresis unit (CE 310 Genetic Analyzer, PE Biosystems). Sample sheets are imported from GMI-LIMS directly into the 310-collection software. Log-data are recorded by 310-collection software. Genescan software is executed by the 310-collection software after the run, and the Genescan files are transferred to the fileserver.

Allele Designation

At this point, two independent operators use Genotyper software for allele designation. We developed specific macro procedures for Genotyper software to simplify and standardize allele designation. During allele designation, the operator logs important information and comments into GMI-LIMS. GMI-LIMS prevents operator 1 from reading the log-data of operator 2 and vice versa. The macro creates data export-files.

Import and Comparison of Results

Genotyper result files are imported into GMI-LIMS. Again, GMI-LIMS checks batch identifier, membership of a sample in a batch, and the DNA profiles of the positive and negative controls. Further, the plausibility of allele values is checked. GMI-LIMS compares the result data of the two independent operators and highlights the differences. In addition, partial profiles or profiles containing rare alleles are highlighted. At this stage a supervisor provided by GMI-LIMS with log-data from operators 1 and 2 can resolve differences or initialize re-PCR or re-extraction procedures (Fig. 2).

Data Transmission

After a checking procedure, GMI-LIMS establishes a modem connection to the national DNA database server system and a profile is loaded into the database. The transmission protocol consists of a check sum calculated by a cyclic redundancy check (CRC)-algorithm.

Duplicate Analysis in Case of Matching Profile

When a match with the National DNA database occurs, GMI-LIMS initializes duplicate analysis of the respective sample. After GMI-LIMS indicates the archive position of the duplicate sample, a DNA-typing procedure follows. The procedure is simi-

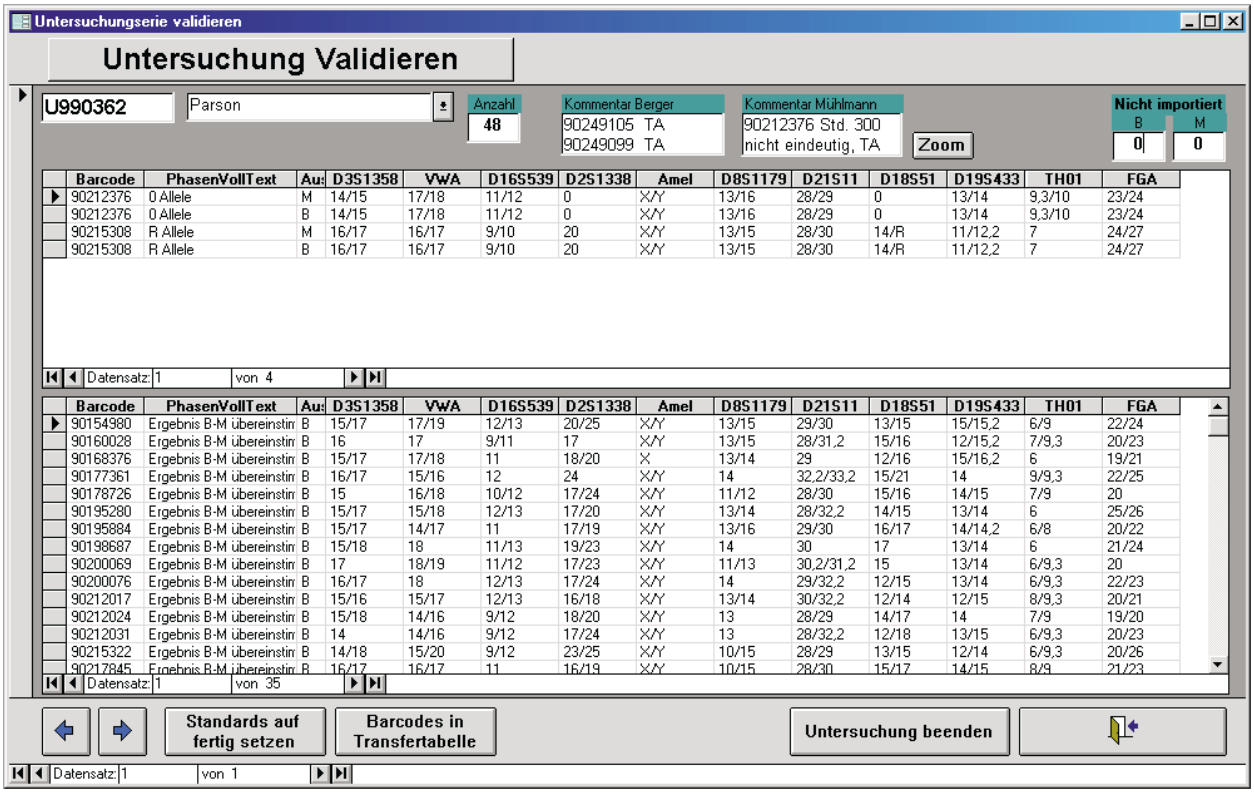

Figure 2. This screen is used by the supervisor to control result-data of operator 1 and 2. The upper window displays short tandem repeat (STR) profiles – one partial profile (D18S51 and D2S1338 dropped out), one with a rare allele (D18S51). In the lower window, STR profiles without differences between the analyses of both operators are displayed. Reprocessing of samples can be managed at this stage.

lar to the first analysis mentioned above. In addition to the normal procedure, GMI-LIMS compares the first profile with the duplicate profile, and then transmits a confirmation report to the national DNA database. The responsible police station is informed about the match only after the profile has been confirmed.

Results and Discussion

Basics of Laboratory Information Management System

It took us about nine months to develop GMI-LIMS and to make it fully operational. We decided to develop GMI-LIMS in the laboratory primarily because it meant more flexibility and could save time. The cost argument, on the other hand, was not decisive. However, the on-site development of GMI-LIMS led to a streamlined reorganization of the laboratory.

Novell Netware 5.0, as a network operation system, enabled a seamless integration of Apple Macintosh-based and Microsoft Windows-based hardware and software platforms.

Microsoft Access 2.0 development tool kit allowed for designing a very smooth user interface and the programming of a made-to-order functionality, and facilitated adressing the above mentioned software packages controling the specific laboratory instruments. This is the basis of a seamlessly integrated application software platform on top of the hardware environment.

Bar Code Technology

The chosen bar code symbols in conjunction with the selected bar code label material and glue proved to be a very robust and reliable system through all the stages of laboratory process. The labels can sustain temperatures between –20^{0}C and +90^{0}C, and the readability is excellent during the whole process. The print format of the "interleaved 2 of 5" code is small enough and so the bar codes can be stuck on 1.5 mL reaction tubes.

The first digit of the 8-digit sample identifier allows sample collectives, such as reference samples, controls, or samples from intelligence screens, to be marked differently. It simplifies the administration and allows the appropriate management of result data.

Scanning a complete batch with the bar code reader takes less then two minutes.

Sample Handling

Sample handling was improved and sample storage optimized by the integrated archive system. The retrieval of a specific sample from storage – we currently store about 35,000 samples – is made easier.

GMI-LIMS guarantees the identification and the defined well position of a specific sample on a specific plate. Complete sample- and batch-histories can be recorded and reconstructed by GMI-LIMS. A very important advantage is that GMI-LIMS relieves staff of administrative work and frees them up to concentrate

on controlling the process on the whole and on data analysis as such. Paperwork has been reduced, whereas necessary documentation, especially for quality assurance purposes, can be generated and printed by GMI-LIMS.

Robotic System

Implementing the robotic system for extraction and PCR set-up by means of software-controlled temperatures, movements, and pipetting led to stable and reproducible laboratory methods. The instruments were programmed so that process safety has priority over time.

Data Analysis and Transmission

The use of the Genotyper macro minimizes both a loss of time and typing errors during allele designation. The automated check of the positive control by GMI-LIMS confirms that the ladder alleles are designated correctly. Software-aided supervision of result data in combination with automated management of reprocessing samples decisively improves the system. Due to automated data transmission of DNA profiles into the national DNA database, additional manual transcription of result data is avoided, and furthermore, the implemented CRC checksum guarantees a correct transmission. Software-controlled management of initializing duplicate sample analysis after obtaining a match makes the process complete.

Conclusion

Since a streamlined workflow has been established, the use of the overall system led to a minimization of manual user input and paperwork. A small team of three operators has been able to process more than 12,000 samples a year by operating two capillary electrophoresis units. Integration of instruments and administration by GMI-LIMS led to a reduction of the turn-around time to about 8 days in average (including the time required for reprocessing of the samples) (5).

The architecture of the system is open insofar that new instruments can be integrated and the sample through-put can be easily upscaled.

Acknowledgment

We thank the staff at the Institute of Legal Medicine for their outstanding work in technical development and their valuable assistance and commitment throughout the entire project. We thank the database staff at the Ministry of the Interior (Abt. II/10) for the excellent cooperation. This study was supported by a grant of the Oesterreichische Nationalbank, Jubiläumsfondprojekt No. S97/4 (Austrian National Bank).

References

1 Parson W, Steinlechner M, Scheithauer R, Schneider PM. National DNA intelligence databases in Europe – report on the current situation. In: Promega Corporation. Proceedings of the Ninth Symposium on Human Identification; 1998; Madison, Wisconsin, USA; p. 52-4.

2 Werret DJ. The national DNA database. Forensic Sci Int 1997;88:33-42.

3 Werret DJ, Sparkes R. 300 matches per week – the effectiveness and future development of DNA intelligence databases. In: Promega Corporation. Proceedings of the Ninth Symposium on Human Identification; 1998; Madison, Wisconsin, USA; p. 55-62.

4 Hopwood A, Brookes J, Shariff A, Cage P, Tatum E, Mirza R, et al. A fully integrated robotic system for high sample throughput within a DNA databasing unit. In: Promega Corporation. Proceedings of the Eighth Symposium on Human Identification; 1997; Madison, Wisconsin, USA; p. 20-4.

5 Parson W, Steinlechner M. Efficient DNA database laboratory strategy for high through-put STR typing of reference samples. Forensic Sci Int. In press 2001.

Received: March 23, 2001

Accepted: April 18, 2001

Correspondence to:

Martin Steinlechner

Institute of Legal Medicine

University of Innsbruck

Muellerstrasse 44

A-6020 Innsbruck, Austria

martin.steinlechner@uibk.ac.at