Status of intellectual protection for the genetic markers in swine breeding

And its potential influence on Norsvin's research and intellectual protection strategy

Konstantin Moiseenko

Høgskolen i **Hedmark**

The Master study program in Applied and Commercial Biotechnology

2BIO101 Master's Thesis

HØGSKOLEN I HEDMARK

2011

Acknowledgement

I am heartily thankful to my supervisor, Dr. Eli Helene Grindflek, whose encouragement, guidance and support from the initial to the final level enabled me to complete this Master Thesis. I would also like to express my gratitude to my Høgskolen Hedmark advisor, Prof. Sigbjørn Lien, for his time and valuable feedback, to Prof. Odd-Arne Olsen who encouraged and challenged me at the beginning of this project and to Marianne H.S. Hansen for her guidance during my laboratory work.

A very special recognition needs to be given to Dr. Maren Moe van Son who went out of her way and invested time in providing me with extensive help and support during my writing.

Lastly, I offer my regards and warmly acknowledge to all my teachers and classmates for their unselfish and unfailing moral and intellectual support.

Table of content

Abstract

Pig is one of the first animals that were domesticated, and pork is one of the most widespread meat products in food market. At present, up-to-date genetic and biotechnological approaches form a solid basis for the modern swine breeding methods. Key United States and European companies broadly use new genetic marker based methods for increasing productivity traits of their populations.

The relevant worldwide patent information about genetic markers in swine breeding was collected during this project. All collected patent documents (145 documents in total and 84 nonrecurring documents) were analyzed in a variety of ways. The present status of intellectual protection together with the last year's and future trends were revealed. General suggestions about Norsvin's research and intellectual protection strategy were done.

Genotyping with markers of interest for a limited number of samples from Norsvin's Landrace and Duroc pig populations was carried out in addition to the main line of the project.

Abbreviations

D	Duroc
EPC	European Patent Convention
EPO	European Patent Office
HWE	Hardy-Weinberg equilibrium
IP	intellectual protection
IPC	International Patent Classification
L	Norwegian Landrace
PCT	Patent Cooperation Treaty
SNP	Single Nucleotide Polymorphism
TRIPS Agreement	Agreement on Trade-Related Aspects of Intellectual Property Rights
USPC	United States Patent Classification
USPTO	United States Patent and Trademark Office
WIPO	World Intellectual Property Organization

Figures

Tables

1. Introduction

1.1 Goals

The main goal of this Master project was to collect relevant worldwide patent information about genetic markers in swine breeding and use this information for further evaluation of status of intellectual protection (IP status) for this field. The sub goal was to elucidate, in a general way, the potential influence of present IP status of marker assisted swine breeding on the Norsvin's research and IP strategy.

To achieve the main goal of the project it was necessary to: 1) choose types of patent documents for which search should be performed, 2) develop search strategy, and 3) make a decision about the way of storing, managing and analyzing the search results.

Additionally it was decided to check with experimental work possible usefulness of some of the most relevant patented markers.

1.2 The ways of achieving the goals

Types of documents included in the search

Patent search was performed within US, EPO and PCT patent documents. Search within these types of documents was preferred because almost all inventors worldwide typically try to get patents through at least one of these patent authorities (Bryant, 1998). Hence searches in these documents are the best way for revealing overall IP status of genetic markers in swine breeding.

Search strategy

Development of the searching strategy is the first and the most important part in a patent search. In a broad term search strategy is the planning of how to look for information. The search strategy should be performed before doing patent search. A well designed search strategy should allow search in many different public patent databases and help finding relevant information in a time saving manner (Hunt, Nguyen, & Rodgers, 2007).

Data storing, management and analyzing

Because more than 100 patents were expected to be obtained during patent search, it was obvious that it was necessary to have some system to simplify organization, storing and retrieving these large amounts of data. A database was developed to achieve this.

Usually simple databases that include just one table can be done using MS Excel electron tables. In fact MS Excel is not a database but it is something more technically known as a spreadsheet for maintaining the data on a limited scale (Fuller, 2007).

It is more comfortable to work with electronic tables if the number of records is low. When number of data increases it is hard to manage them. Usually the reason for this is bad structure of the data. Also computer operation speed decreases with increasing number of data (Groh, 2007).

MS Access helps storing large amounts of data in one relational database and managing them using forms, queries and reports. This saves some computer memory, increases the data processing rate and helps to avoid mistakes and duplication (Lambert & Lambert, 2007). MS Access was therefore chosen for the data storing and managing.

Experiments

The decision was made to carry out genotyping with markers of interest for a limited number of samples from Norsvin's Landrace and Duroc pig populations, to determine if the markers are segregating in the populations or not and whether they are in the Hardy-Weinberg equilibrium (HWE). Although presence of polymorphisms for some claimed markers and knowledge about their HWE are not straight forward indication of the markers' utility obtaining this information is clarifying potential usefulness of these markers in future association studies.

2. Background

2.1 Genetic markers in swine breeding

Pig is one of the first animals that were domesticated, and pork is one of the most widespread meat products in food market. At present, up-to-date genetic and biotechnological approaches form a solid basis for the modern swine breeding methods.

Although the traditional methods of selection are generally recognized, and has been effective during a long period of time, it is now obvious that using just these methods cannot provide proper efficiency level of breeding work (Rothschild, Stalde, & Dekkers, 2010).

It is very simple to select animals for Mendelian inherited traits (examples of Mendelian inheritance include autosomal dominant, autosomal recessive, and sex-linked genes) using traditional methods of selection (Lambe & Simm, 2004). In contrast, selection of organisms which bear some advantageous quantitative traits (traits that are controlled by multiple genes and highly influenced by environment) is a very hard task. This selection typically should be carried out for generations. Marker-assisted selection (MAS) provides a great opportunity for accelerating this procedure (Dekkers, 2004). MAS is a process whereby a genetic marker is used for indirect selection of a trait of interest.

Broadly speaking, a genetic marker is a nucleotide sequence variation which location is precisely defined on the chromosome or on the part of the chromosome (Navajas & Simm, 2004). Genetic markers appear in result of different types of DNA mutations: substitution mutations (point mutations/Single Nucleotide Polymorphisms (SNPs)), rearrangements (insertions or deletions) and errors in replication of tandem repeated DNA (Beuzen, Stear, & Chang, 2000).

Genetic markers are usually located within the non-coding regions of the chromosomes and they are therefore usually not directly affecting the traits and consequently often selectively neutral (Beuzen, Stear, & Chang, 2000). However, the vast majority of genetic markers are placed in the close proximity to the genes that provide development of some traits and can be used like "tags" for these genes. Even if genetic markers are located within a gene they generally do not have any straight influence on the gene's function, they are just linked with some features that provide this function and can be utilized like "tags" for allelic variations of a gene (Beuzen, Stear, & Chang, 2000).

Genetic markers can be classified in variety of ways. First, they can be classified according to the methods of detection, like restriction fragment length polymorphism (RFLP) or polymerase chain reaction restriction fragment length polymorphism (PCR-RFL) markers, random amplified polymorphic DNA (RAPD) markers, amplified fragment length polymorphism (AFLP) markers and so on. Second, they can be classified according to their origin into simple sequence repeats (SSRs) or "microsatellites" markers, SNP markers and so on. Third, genetic markers can be divided into dominant and co-dominant markers. This division is based on number of forms (so called marker alleles) in which the marker can be detected. Dominant markers can be either present or absent (i.e. can study only one out of minimum two alleles) whereas for co-dominant markers entire allelic variations of their locus can be distinguished (i.e. can study two or more alleles for each locus).

Identification of genes, which allelic variances change predisposition of an organism to have some trait of interest is a great tool that allows evaluating genetic profiles of organisms and using these profiles for further breeding purposes. That is why key US and European companies broadly use new genetic marker based methods for increasing productivity traits of their populations.

2.2 Patents

2.2.1 General information about patents

A patent is a government-granted monopoly on an invention to an inventor or their assignee for a limited period of time in exchange for a public disclosure of an invention (U.S. Department of Commerce, 2008). The term "limited period of time", means the maximum period of time where a patent can be maintained into force. It is usually expressed in number of years (typically up to twenty years) either starting from the filing date of the patent application or from the date of grant of the patent (Engelfriet, Frequently Asked Questions: General questions, 2002). As long as the patent is valid the owner has to pay a yearly fee in order to keep the patent in force. Otherwise the patent will be lapsed before its term (U.S. Department of Commerce, 2008).

The scope of a patent is information about what the patent exactly is protecting. By rule all this information should be included in a patent's claims (Rothschild & Newman, 2002). All patent's claims are organized in a single set and a lot of standard sets of claims can be found in the patent literature (Bryant, 1998). The terms which are used in the claims can be defined during the whole patent document. These arbitrary defining terms can significantly broaden the scope of a patent (its literal wording) (U.S. Department of Commerce, 2008).

Claims are used to define the validity of a patent by comparing them against the prior art and to make a decision about infringement of exclusive rights which is given through the patent protection. To infringe the patent, infringing product must contain each and every element of claims (Engelfriet, 2005). As soon as inventors try to cover as much as possible by their patent, in each field of technology specific for this field broadening languages were developed. These languages allow broadening out the scope of the patent beyond its literal wording (Cambia).

Patents are exclusively national affairs, so a patent that was issued in one country cannot provide protection of the invention in any other countries in the world. However, some countries have concluded treaties under which patents can be granted, and these patents automatically become valid in all treaty members (Engelfriet, Frequently Asked Questions: General questions, 2002).

2.2.2 Main categories of patents

International applications

International applications are the patent applications that are filed under the Patent Cooperation Treaty (PCT) which was administered by World Intellectual Property Organization (WIPO) (Patent Cooperation Treaty, 1970).

Two things should be noted about international applications. First of all, these applications do not give any international protection to the applicant. The applicants are free to file whatever they wants, so it can easily happen that the applicants file something that is useless, long known or totally trivial (Engelfriet, Frequently Asked Questions: General questions, 2002). Second, WIPO does not get any responsibility for the issuing patents in any of the PCT countries. It just offers a way to file one application centrally. The applicants then have to enter national phase (foreign filings) of the countries of their choice within the time limits to do so and defend their claims to get a patent in those national offices. That is why PCT applications unlike most patent applications can never result in a patent (Rothschild & Newman, 2002).

However, many inventors all over the world try to get this type of application at the beginning, because PCT applications provide several benefits to applicants. First, it allows the applicant to delay the expenses of filing applications around the world. Second, just one literature search will be performed and all national patent offices indicated on the front page of the publication will use results of this search for document examination. Third, by filing a PCT application, the applicants can postpone the decision on patenting their invention for 30 months (rather than 12 months under the Paris Convention) in the others PCT countries. It allows the inventors more time to assess the commercial viability of their invention (Bryant, 1998).

European patents and applications

The term European patent is used to refer to patents granted under the European Patent Convention (EPC) (European Patent Convention, 1973).

Two ways may be appropriate for getting a European patent. First of all, the inventor can file patent application directly with the country, or countries, in the European Union (EU) in which patents are desired. Second, the inventor can do single filing of a patent application with the European Patent Office (EPO) (Bryant, 1998).

EPO is the official organization which actual legislative power proceeds from EPC. The EPO is not a body of the EU and patent granted by the EPO does not lead to a single EU-wide patent (Rothschild & Newman, 2002). However, the EPO offers a way to file a single patent application which can lead to patent coverage in all the European countries that belong to the EPC. In fact EPO patents get a legislation power through independent national patents in the EPC countries of choice (Durham, 2004).

Usually if the inventor wants to get a patent only in one or two countries of Europe, it may be cheaper to apply directly into national offices of these countries. On the other hand, if the patent is planned to cover all or the biggest part of the EPC countries, then it will be favourable to file directly with the EPO (Rothschild & Newman, 2002).

United States patents and applications

United States (US) patents and applications can be defined as patents and applications granted under the US patent law. The provisions of the law are laid out in Title 35 of the US Code (U.S. Code Title 35) and give authority for the US Patent and Trademark Office (USPTO) (U.S. Department of Commerce, 2008).

US patents give a monopoly right in the whole territory of the US. According to this nobody but the patent holder can make, use or sell the invention in the US (Bryant, 1998). Because the US is a huge market and US patenting system is one of the oldest systems in the world, all inventors are typically trying to get a US patent even if they have other patents all over the world.

2.3 Patenting of genetic inventions

2.3.1 Patentability

In principle all genetic inventions fall into the category of biotechnological inventions. These include inventions which are biological, microbiological, genetic engineering, medical, and agriculture (Lakshmikumaran, 2007). Article 27(1) of the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement) clearly states that patents should be granted for inventions in any field of technology without discrimination, subject to certain clauses (TRIPS Agreement, 1994). This implies that biotechnological inventions and consequently all genetic inventions are patentable subject matter.

The US Courts and EPO have granted patents to genetic inventions. However, the issue of patentability of genes and gene sequences is yet not clear (Jauhar & Narnaulia, 2010). In the EPC Countries the issue can be clarified using Rule 27 of the EPC (Patentable biotechnological inventions). This rule states that all DNA sequences and isolated genes are patentable subject matters as soon as they were isolated from its natural environment and characterized at the same way like chemical compounds (European Patent Convention, 1973). However, at the present stage this rule is narrowed by Biotechnology Directive (98/44/EC). After this directive had been settled it became impossible to get a patent on DNA sequence or isolated gene without any characterization of its function or/and without any idea of its use (Biotechnology Directive, 1998).

In practice the US law is very similar to the European law in the definition of what should be patentable within the area of genetic inventions. While the USPTO often makes the determination as to whether an invention is patentable or not, the standard is actually set by the Courts and is only exercised by the USPTO. The Courts can overrule the USPTO in litigation (U.S. Department of Commerce, 2008). Using the experience of previous US court cases on DNA patents it can be concluded that it is improbable that the US patent law will favour the objection that isolated genes must be categorized as unpatentable discoveries if its actual use was evident or studied by the inventor (Crespi, 2000).

In summary, it can be concluded that to get a DNA patent in US or Europe, broadly speaking, inventors must isolate and identify novel genetic sequences, specify the products of these sequences and specify how these products functions in nature (i.e. its use) (Crespi, 2000). However, some more specific questions related to this field can still occur and should be clarified. Some of these questions are answered below.

2.3.2 General questions

Can different inventors patent the same DNA strands?

Yes they can. Sometimes, different inventors can get a patent on the same DNA strands simply because the strands were discovered using different methods. Also, the methods themselves can be patented (John, 1998).

How will the issuance of a patent on DNA fragments of a gene affect the patenting of full-length genes?

Patent claims limited in scope to specific novel and non-obvious DNA fragments, for example SNPs or expressed sequence tags (ESTs), will not necessarily prevent the future patenting of the corresponding full-length gene of known function, if significant amount of information about the gene and/or protein (which can be the product of a gene) will be disclosed in another patent application (John, 1998).

Can patent be granted on something within a patented sequence/gene?

If the patent on the invention is granted, this patent covers all possible uses of this invention. It means that nobody can use claimed invention in any ways, even if these ways was not mentioned in the claims, without infringing the patent (Engelfriet, 2002). However, method of using the claimed invention may be patentable itself, assuming that this method is non-obvious. In this situation the first patent (on the invention) will dominate over the second patent (on the new method of use), because for practicing the second patent it will be necessary to obtain license on the first one (Rothschild & Newman, 2002). Moreover in some cases fees have to be paid just for examining the claimed sequence or having the ability to invent something new within the sequence that was already claimed. Sometimes the situation can become really complicated when a researcher have to purchase 3-8 licenses at once as a result of allowing multiple patents on different parts of the same genome sequence (John, 1998).

Genetic Testing – Infringement?

This question is two parts:

1. If a laboratory works with the patented isolated genetic sequence, will it infringe the patent?

2. If a laboratory works using the patented method of diagnosis, will it infringe the patent?

There are two possible answers, depending on the examination of the actual granted patent claims. First of all, there may be opportunities for inventing around the patent. It is usually permitted when the patent is old and new laboratory or diagnostic methods have been developed (new sequencing technologies, for example) (Hawkins, 2010). Second, if a valid gene patent exists, and infringing can be proved, inventor can use some legal defenses. The most relevant defense for the purpose of gene patents and diagnostic testing is the experimental use defence. Researches that aim either to verify claimed functions or to determine new functions of a patented gene sequence, to find new genetic markers within a patented gene sequence, to find new information related to a gene sequence and to develop a new genetic diagnostic test will most probably be covered by the experimental use defence (Hawkins, 2010). However it is very important to note that the research tools which has been validated and becomes routine at the time when researcher perform it (like Polymerase Chain Reaction (PCR)) will not be an act done for experimental purposes relating to the subject matter and will not fall within the research exception (Hawkins, 2010).

2.4 Survey of MS Access database management system

2.4.1 General information about data storage and managing

2.4.1.1   Relational data model

A data model is an abstract model which explains the way of data storing and access. There are many possible data models that can be applied for the databases construction. For example:

- Flat file model - data is organized into two-dimensional array of data elements.

- Hierarchical model - data is organized into a tree-like structure.

- Network model - data is organized into network-like structure, which is not restricted to be a hierarchy or lattice.

- Relation model - data is organized using concept of objects and relations between object.

Today almost all databases (including MS Access) correspond to the relational data model (Groh, 2007). A database that is constructed using relations is called relation database. This type of database is described in detail below.

Two-dimensional tables are one of the most natural ways of data representation. Data about different types of objects (object's attributes) is stored inside the different tables in the form of columns and rows and one complete set of data for one object is called a record (Fuller, 2007).

Relation between objects can also be represented in the flat table (Lambert & Lambert, 2007). For instance relation between two tables (i.e. two objects) can be settled by a third table which should store identification number (ID) of records from the first table and ID of records from the second table in the first column and in the second column, respectively (Figure 1).

It can be three types of relations between tables: one-to-one – one record from table 1 is linked with just one record from table 2; one-to-many – one record from table 1 is linked with many records from table 2; many-to-many – many records from table 1 are linked with many records from table 2. In the case of one-to-many relation, table "one" is called main table and table "many" is called slave table (Lambert & Lambert, 2007).

It should be noted that each table of the relational database can be defined like relation, because inside each table, attributes of each record (i.e. intercrosses between record's row and columns) are linked together in the one-by-one manner (Groh, 2007). In the case of the relation database terms "relation" and "table" can be used like synonymous.

Hence all data can be stored and represented using flat tables (Lambert & Lambert, 2007). Each row of each table consists of data about one object. Row is typically named like records and columns like fields of the record. Fields consist of attributes of the record's objects. All records have identical fields, which include different values of attributes. Each field has a strictly defined data type (text, number, date and so on) (Jennings, 2007).

Fields which form a unique identifier in a one-to-one correspondence manner for each database record are called key fields (the same with ID). Key fields are used like a table entry to link tables together and help to perform fast search of data for further representation in the form of queries, forms and reports (Groh, 2007).
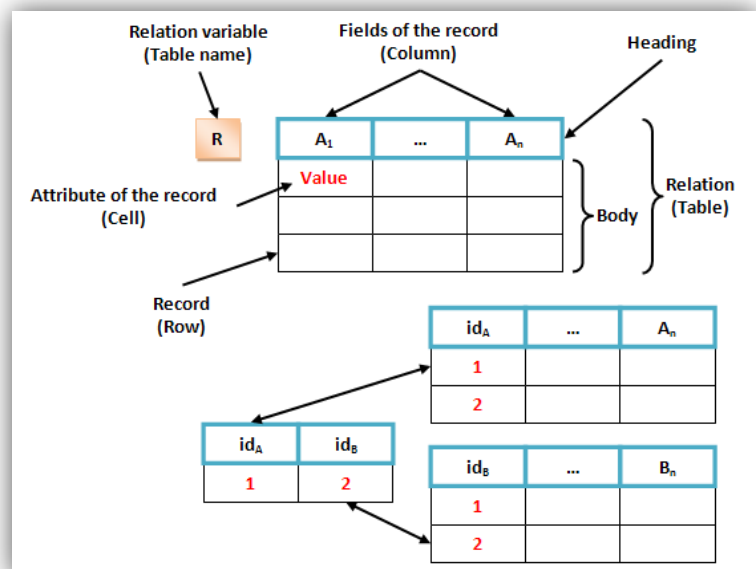
Figure 1. Relational model concepts

2.4.1.2 Database management system

It should be noted that the term database includes just a subset of the tables. All the rest, like storage and maintenance of database's content and data creation and search belong to the database management system (Groh, 2007). Database management system is the specialized program package that is used for working with the data.

The most important function of the database management system is to provide database integrity. It means setting some rules on the relations between tables: It should be impossible to put attribute that do not exist in the main table into the fields of a slave table; It should be impossible to delete attributes from the main table if there are existing links to the records in the slave tables; It should be impossible to change key fields in the main table if there are existing links to the records in the slave tables (Lambert & Lambert, 2007).

All operations with the data typically require using special Structured Query Language (SQL). The main advantage of MS Access database management system is that the main operation with data can be done using visual construction method of queries to the database (Lambert & Lambert, 2007). In this situation SQL query is automatically generated by MS Access.

2.4.2 Main elements of MS Access database management system

Tables

Tables are the main part of each relational database. First, all data within the database are stored inside the tables. Second, tables are storing database structure (fields and their types and properties) (Groh, 2007).

Queries

Queries are used for extracting data from the tables and representing them in a user friendly manner. All operations of data searching, selection, sorting and filtration are done using queries. Queries allows to do changes in the data using a given algorithm, creating new tables, doing automatically fill up of the tables, importing data from other sources, doing main calculation inside tables and so on (Groh, 2007).

Forms

If queries are specialized tools for data selection and analysis, forms are the tools for data input. Forms have the same idea, allowing the users to fill up just fields that they should fill up. For this purpose special control elements (buttons, checkboxes and so on) can be located on a form which provides more comfortable input and navigation (Groh, 2007).

Reports

In the term of structure and properties reports are very similar to forms, but their purpose is just to represent data, typically in the printed format. Hence reports have special features for data sorting, grouping and data representation design (Groh, 2007).

2.5 Patent search

2.5.1 Introduction to the search strategy development

2.5.1.1   Main information

For finding relevant information it is really important to get a focus on what has been previously claimed or could be potentially claimed in the area of interest. Think about what should be found by concentrating upon the idea of the inventions. In other words, to find relevant information it is necessary to have the key concepts of a search (Miller, 1999).

Key concept is the written equivalent of the idea of the invention. Key concept should be written down in a form of an affirmative sentence using ordinary writing language (Calishain, 2004). This sentence will be further braced into key terms (keywords and key phrases) which should be used for generating queries to the online patent databases (Figure 2). Number of key terms can be raised by using synonyms, spelling variations and so on (Baylin & Gill, 2005).

Queries to the online patent databases are preferably formed from keywords and key phrases using Boolean operators. A query can be broadened by using OR operator instead of AND operator for joining terms together, using more general search terms and using truncations (Calishain, 2004).

When a query is constructed, the searcher can go to the on-line patent storages to perform preliminary patent search. Performing search in the right places is the main requirement for the success (Hunt, Nguyen, & Rodgers, 2007). For finding a good place of patent search the searcher should ask himself the following questions: Are there some places that collect patents, which directly relate to my searching field? How comprehensive is the patent collection in this place of search? How much searching facilities are available if I perform a search at this place? And, how user friendly is search results' representation in this place of search?
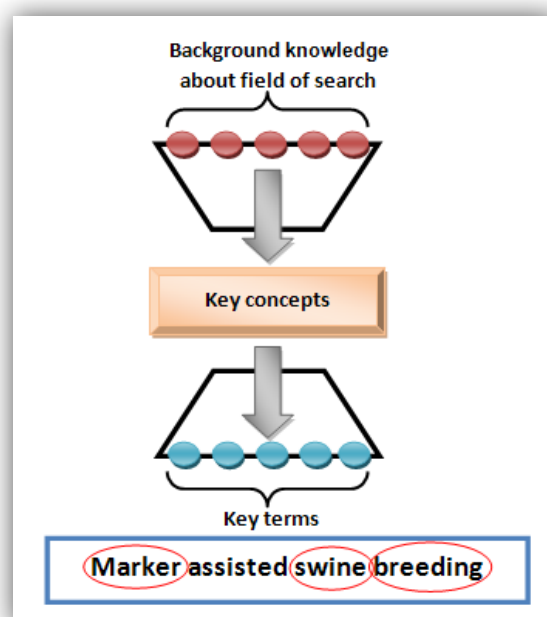
Figure 2. Idea of key concepts

After preliminary search have been performed, it can be useful to do some adjustments to the search strategy. This first feedback can help to refine queries, to find new searching criteria and to organize the search into one strain line in the most comfortable and time saving manner (Hunt, Nguyen, & Rodgers, 2007).

2.5.1.2 Using patent classification for finding additional search criteria

Patent classification is a system for organizing patents by subject matter. Main purpose of this system is to facilitate patents retrieval by manual patent searches. Two main patent classification systems: United States Patent Classification (USPC) and International Patent Classification (IPC) are described in details below.

The USPC is an official patent classification system used and maintained by the USPTO. It contains over 400 classes. Each class is identified by a class number and have a title that describes its subject matter. Each class is subdivided into a number of subclasses. Each subclass also has a subclass number and descriptive title. The subclass number may be an integral number or may contain a decimal portion and/or alpha characters. A complete identification of a subclass requires both the class and subclass number and any alpha or decimal designations; e.g., 435/6 identifies Class 435, Subclass 6 (Figure 3) (Hunt, Nguyen, & Rodgers, 2007).

The IPC is a hierarchical patent classification system created under the Strasbourg Agreement (Strasbourg Agreement, 1971). The Strasbourg Agreement is one of a number of treaties administered by the WIPO. Each classification term consists of a symbol such as C12Q 1/68. The first letter is the "section symbol". This is followed by a two digit number to give a "class symbol". The final letter makes up the "subclass". The subclass is then followed by a 1 to 3 digit "group" number, an oblique stroke and a number of at least two digits representing a "main group" or "subgroup" (Figure 3) (Hunt, Nguyen, & Rodgers, 2007).
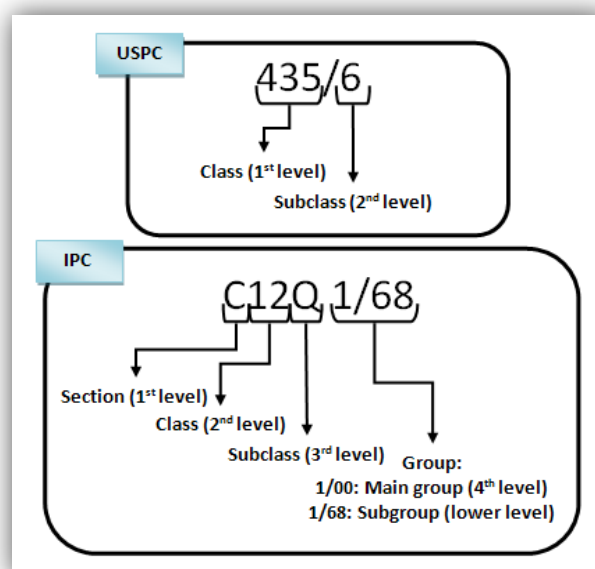
Figure 3. Patent classifications

IPC is used by almost all patent authority all over the world. However, vast amount of patents that were issued in the US had not been satisfactory classified according to IPC. The reason for this is, first of all, that US patent examiners classify patents with US marks more accurately than they do with IPC marks and secondly, this classification was established 100 years later than UPC (Hunt, Nguyen, & Rodgers, 2007).

2.5.2 Searching public databases

2.5.2.1   Types of databases

There are two types of patent databases that should be pointed out: primary patent databases and secondary patent databases. Primary patent databases are storages of data deriving directly from issuing patent authority. By definition these databases contain all information about all issuing patents (Baylin & Gill, 2005). Secondary patent databases are storages of data that was extracted from primary patent databases and organized in a different way (Baylin & Gill, 2005). Secondary patent databases are usually incorporated into a search engine (Figure 4).

Search engine is a suite of programs which contain: "Spider" - program that upload Web-pages from the Internet into a search engine; "Crawler" - program that go through uploaded Web-pages and search for a links to other Web-pages, thereby giving direction of further movement to the Spider; "Indexer" - program that divide Web-page into distinct parts and index this parts; "Database" - program that store all indexed data; "Result engine" - program that analyze query from user, perform a search in the incorporated database and return results that can be relevant to the query (Figure 4) (Calishain, 2004).

Figure 4. Primary and secondary database concepts

In a simple term the work of search engine can be described as follows: Search robot (Spider plus Crawler) go through all relevant web-sites and upload all relevant web-pages from these web-sites into the search engine. Inside the search engine web-pages are indexed and saved in an incorporated database by indexer. After the user puts a query into the result engine, this program analyzes it and returns results from an incorporated dat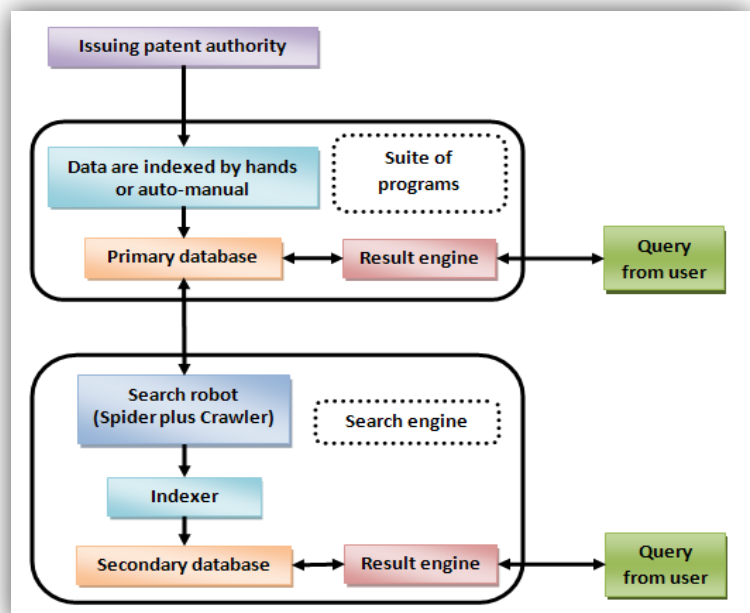abase that possibly solve the query. Therefore the user works only with one part of the search engine, namely, the result engine and can get the results from search engine incorporated database only. So results can be not comprehensive since they were already filtered by search robot (Miller, 1999).

Primary patent databases are also incorporated into a specific suite of programs, but in contrast with the search engine this suite contains just the result engine and the database themselves. All data are indexed by hands or auto-manual. These databases contain all patents that relate to their field and all comprehensive information about each patent (Figure 4) (Calishain, 2004).

Both primary and secondary databases have their own pros and cons. Secondary databases usually have more user friendly search interface, collect specific information about one field from many primary databases and allow doing worldwide search. However, no secondary databases can give a guaranty that they contain all patents that was issued by patent authorities. Only primary databases can give this guaranty (Hunt, Nguyen, & Rodgers, 2007).

2.5.2.2   Types of search

By types of search means types and formats of queries that are acceptable for the result engine. All result engines are using specific internal languages. Basically these languages have specific syntactical rules for distinguishing key words from key phrases, for combining key terms together and for distinguishing filed within patent document where user suppose to perform a search using each of the key terms.

Usually the result engines accept several types of query input (Miller, 1999). The user can either type query directly on internal language or use adopting interactive forms, information from which will be further automatically analyzed and rewritten on internal language by the result engine.

All searches can be classified according to the number of fields within a patent document where the user wants to perform search into simple search (just one field) and advanced search (many fields), and according to the format of query input into Boolean search (using internal language) and structural search (using interactive forms) (Hunt, Nguyen, & Rodgers, 2007).

Typically all patents web-sites allow the user to do simple structural search (often called just simple search or quick search), advanced Boolean search (often called Boolean search) and advanced structural search (often called either structural search or advanced search).

3. Materials and methods

3.1 Patent-oriented MS Access database construction

In the MS Access, like in the other types of database management systems, data structure should be constructed before actually starting database creation. Therefore, initial phase of the database construction should be done on a paper and thereafter tables, forms and other elements can be done using the program. Good data structure is the main requirement for creation of effective and user friendly databases. It should be noted that during database testing more and more adjustments are typically required and if data structure was done well it is not a problem to change database without starting from scratch.

Database construction was done in four stages:

1. The main information that should be stored and the way this information should be managed were defined. In other words, the purpose of the database was specified.

2. General list of fields that should be relevant to the database purpose was created. At the start it was 24 fields which then were filtered during database improvement.

3. All fields were combined into groups using functional criterion. It was done in the form of entity-relationship model (ERM). Entity-relationship modelling is a database modelling method used to produce a type of conceptual diagram of a system (entity-relationship diagram, ER diagram, or ERD). In a form of ERD system is described using linked blocks, which represent entities (objects), entity's attributes and relations between entities (Groh, 2007).

4. Relational data model and database elements were created using MS Access Database management system.

First stage was already discussed in the Introduction part and will not be described below. Complete list of fields (after filtering) is presented in Table 1 and Table 2. ERM of the patent-oriented database is depicted in Figure 5. Results from final (fourth) stage are presented in the Results section.

Table 1. List of fields for the patent-oriented database

Item		Description
Bibliographical data	**Patent/Application Number**	This item includes: • **US Patent number** • **US Application Document number** • **WIPO Publication number** • **EP Publication number**
	Type of the Document	This item includes: • **International application** • **European patent/application** • **US patent/application**
	Title of the invention	This items include **standard bibliographical data** from the front page of a patent
	Date of Issue	
	Language of the Document	
	Primary Inventor	
	Abstract	
	Claims	
	Primary USPC class/subclass	
	Primary IPC class/subclass	
	Organizations	This item includes all **organizations** (Companies, Universities, National Departments and so on) that mentioned **on a front page** of a Patent (i.e. Applicants for the International application, Assignees for the US patent/application)
	Links to the proper Public Databases	This items include **links to the public patents databases**
	Also published as	This item includes **references to the similar patents/applications** from other patents authorities

Table 2. List of fields for the patent-oriented database (continue)

Item		Description
Additional data	**Internal Database Number**	This item includes **number of unique document** in Internal MS Access database (since some patents/applications from different patent authorities can have the same content they should be grouped together with a single document's number)
	Internal Data Base classification	This item includes: • **Exterior** • **Health** • **Meat Quality** • **Production** • **Reproduction**
	Status in Norway	This item includes: • **The same patent is granted** • **The same patent can be granted** • **Not granted or cannot be granted** • **Granted** • **Can be granted**
	Breeds	This item includes all **breeds** that where mentioned in **Example/Experimental part** of a patent/application.
	Type of Marker	This item includes information about **types of markers** that were mentioned in the **example/experimental part** of a patent/application. However it does not mean that other types of markers were not covered by that patent/application (In the patent/application can be noticed that other markers can be suitable for the same genotyping).
	Genes	This item includes information about **genes** that was mentioned in the patent, including their **name and common abbreviation.**
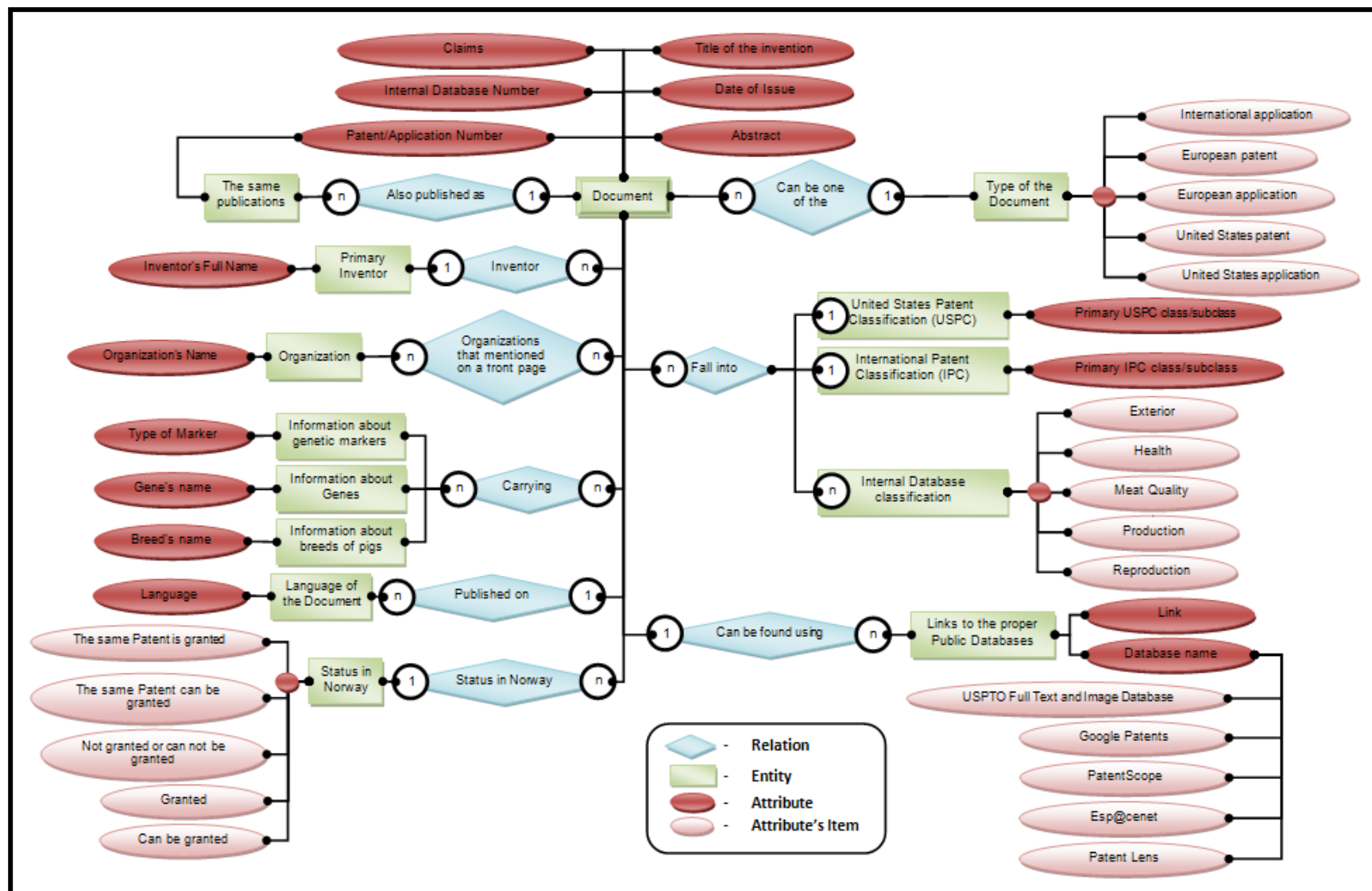
Figure 5. Entity-relationship diagram of the patent-oriented database

3.2 Patent search

As was already mentioned search strategy development should include eight stages:

- Figure out the main focus of a search using information that is already known and better judgments about information that should be found.

- Identifying and defining the key concept of a search.

- Extracting key terms.

- Finding alternative key terms: synonyms (boar – pig), spelling variations, variations of a root word (pig – piglet), plural/singular forms (marker – markers), acronyms (ESR gene – Estrogen receptor gene).

- Generating queries and performing preliminary patent search using proper public databases.

- Refining the preliminary search.

- Analyzing results of preliminary search. Finding additional search criteria.

- Final patent search.

Realization of these stages is presented below.

3.2.1 Identifying and defining key concepts and terms

To define the main focus and the key search concepts, all relevant patents (20 patents) found by previous searchers were examined. General queries (key concepts) were constructed, using information from this examination.

General queries:

- Porcine polymorphisms and methods for detecting them
- Marker assisted swine breeding
- Approaches to identify genetic traits in animals
- Genetic marker based pig selection

Then general queries were broken down into its key terms and alternative terms were found.

All key terms fell into three groups (Table 3):

- Main field – this group include key terms that relate to the main field of a search.
- Species – this group include key terms that relate to the species of interest.
- Traits – this group include key terms that relate to the traits of interest.

Table 3. Key terms (key phrases in the quotation marks can be divided into key words)

Main field	Species	Traits
• gene/genes • QTL/QTLs • "genetic polymorphism/polymorphisms" • "genetic marker/markers" • "molecular marker/markers" • "DNA polymorphism/polymorphisms" • "DNA marker/markers" • "marker assistant selection" • "marker assistant breeding"	• pig/pigs • piglet/piglets • swine • boar/boars • porcine • pork • sow/sows	• fatness • "animal growth" • "meat quality" • "feed efficiency" • "reproductive efficiency" • "disease resistance" • "carcass traits" • "litter size" • "boar taint" • "weight gain" • "muscle growth"

3.2.2 Survey of publicly available databases used during patent search

3.2.2.1   General information about databases

United States patents and applications

USPTO Full-Text and Image Database

Main information: It is the official site of the USPTO. The database is normally updated every Tuesday.

Searchable documents: Front page data (with revised USPC), text of claims and description of US patents from 1976 are available. Images of patents from 1790 are available. Published applications from 15 March 2001 are available as a separate database in the same way as grants (the "assignees" are often not given). Issued and Published Sequences are available as a separate database. Revised versions of patents following litigation and correction slips are not included.

Search facilities: quick search, advanced Boolean search and number search are available for both patents and applications. Published sequences are searchable by document number and date range.

Google Patents Search

Main information: It is a search engine from Google that index patents and patent applications from the USPTO.

Searchable documents: Approximately 7 million full-text US patents and over a million US patent applications are available.

Search facilities: quick search, advanced Boolean search and advanced structural search are available.

DNA Patent Database (DPD)

Main information: It is a Search engine that index DNA-based patents and patent applications from the USPTO. The database is normally updated weekly.

Searchable documents: Over 57,000 links to the DNA-based patents issued from 1971 to the present and over 89,000 links to the DNA-based patent applications published from 2001 to the present.

Search facilities: quick search, advanced Boolean search and advanced structural search are available.

World wide search

PatentScope

Description: It is the official site of the PCT. It is updated each Thursday.

Searchable documents: Contains around 1.8 million published International Patent Applications and the collections of patents from some national phases.

Search facilities: quick search, advanced Boolean search and advanced structural search are available.

Esp@cenet

Description: This is the EPO gateway. The database is normally updated weekly.

Searchable documents: Contains granted European Patent which indicated by a "B" document kind code and European published applications which indicated by an "A" document kind code. The "worldwide" format enables searching across a vast amount of national phases data.

Search facilities: quick search and advanced structural search are available.

Patent Lens

Description: It is a Search engine that index patents and patent applications from PCT, USPTO, EPO and Australian Patent Office (AU). The database is normally updated weekly.

Searchable documents: Contains full-text of over 8 million patents and applications from PCT, USPTO, EPO and AU with list of issued and published sequences.

Search facilities: quick search, advanced Boolean search and advanced structural search are available.

3.2.2.2   Databases comparison

Brief database comparison is showed in the table bellow (Table 4).

Table 4. Database comparison

	United States patents and applications			World wide search		
	USPTO Full-Text and Image Database	Google Patents Search	DNA Patent Database (DPD)	PatentScope	Esp@cenet	Patent Lens
Type of database	Primary	Secondary	Secondary	Primary just for IPA	Primary just for EP	Secondary
Full-text documents	Yes	Yeas	No just links to USPTO	Yes	Yes	Yes
Sequences search	Yes	No	No	No	No	Yes
Search facilities Simple search	Yes	Yes	Yes	Yes	Yes	Yes
Boolean search	Yes	Yes	Yes	Yes	No	Yes
Structural search	No	Yes	Yes	Yes	Yes	Yes
Data representation interface	Medium	Good	Medium	Excellent	Good	Excellent
Additional features	No	No	No	Result statistics	Patent family, IPR status	Patent family, IPR status

3.2.3 Refining preliminary search

3.2.3.1 Query optimization

Virtually all of the key terms that were mentioned previously and all of their possible combinations are suitable for the construction of queries. However the relevance of search results is not equal. By relevance means how much useful patents can be extracted from the public patent database using this or that query. Relevance can be increased on the one hand by successful gathering of key terms into groups using Boolean operators and on the other hand by performing search into those fields of the document where density of this groups of terms should be maximal.

The most successful queries were produced by combination of terms in the same group (Table 3) with Boolean operator "OR" and gathering this "OR – tuples" with Boolean operator "AND". These "OR+AND – tuples" were applied for the search within both "Abstract (Front page)" and "Any field" of documents at the same time using advanced search function. Three of the most successful ways of key terms gathering and searching fields choosing are depicted on the diagram below (Figure 6).
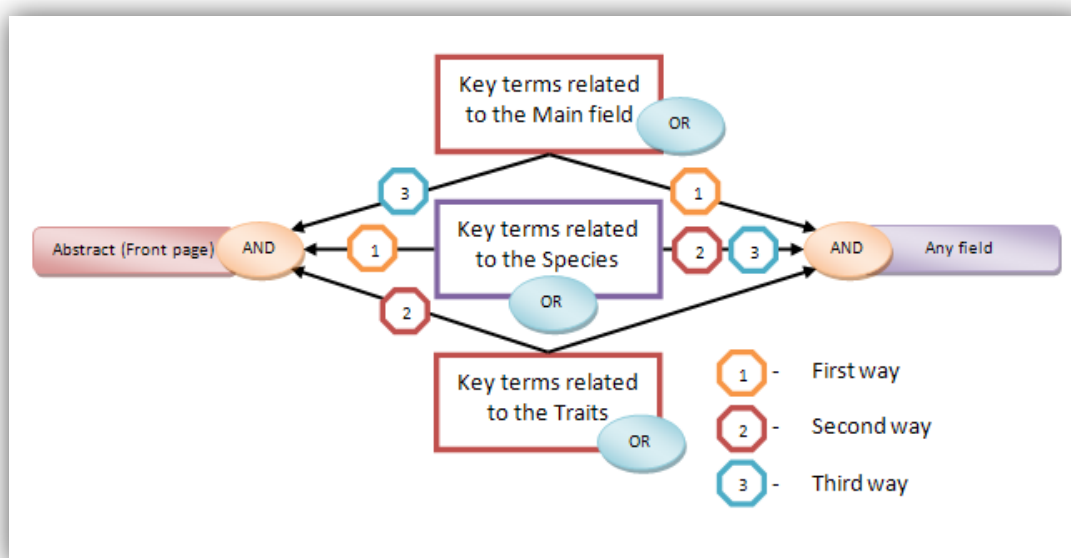
Figure 6. Queries construction

3.2.3.2 Finding proper search class

All patent documents that were collected during preliminary search (70 documents) were analyzed with the purpose to find proper patent class which can narrow down searching area. Main attention was given to the primary class. Primary class is the main class in which patents should to fall according to the opinion of experts in the patent authority.

Results of patent analyzes is showed in Figure 7. It should be noted that almost all patent documents for which classes 435/6 and C12Q 1/68 are not marked like a primary classes steal fall into these classes like into "cross-referenced" classes.
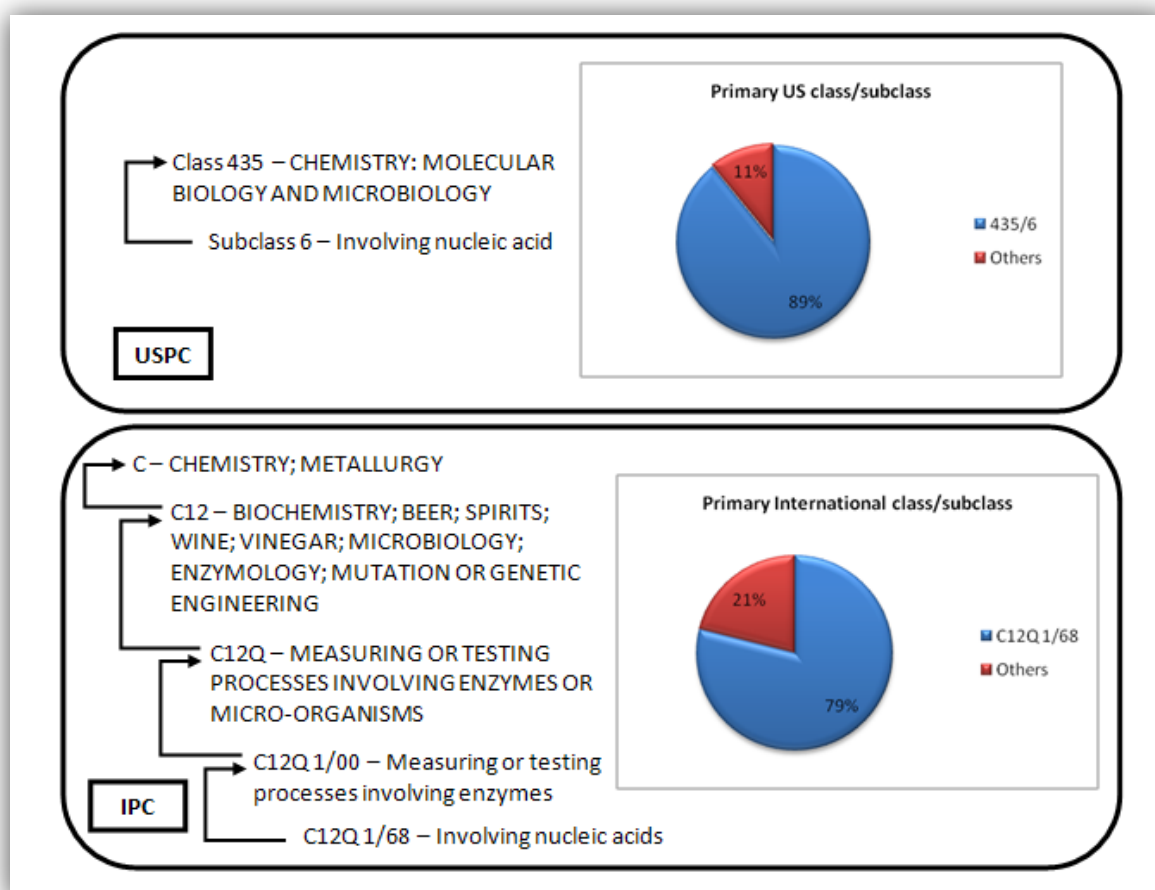
Figure 7. The most common primary USPC and IPC classes

3.3 Genotyping genetic markers of interest

3.3.1 Main information

The iPLEX Gold reaction kit (Figure 8) was used for genotyping (SEQUENOM, 2009).

Short description of the steps for the iPLEX Gold assay is presented below:

1. Isolation and amplification of the genomic DNA samples
2. Neutralization of unincorporated dNTPs in amplification products using shrimp alkaline phosphatase (SAP).
3. Performing the iPLEX Gold reaction, this involves the enzymatic addition of terminator nucleotides into the diagnostic site.
4. Transfer of the iPLEX Gold reaction's products onto a SpectroCHIP array. The SpectroCHIP array is then analyzed by the MassARRAY analyzer.
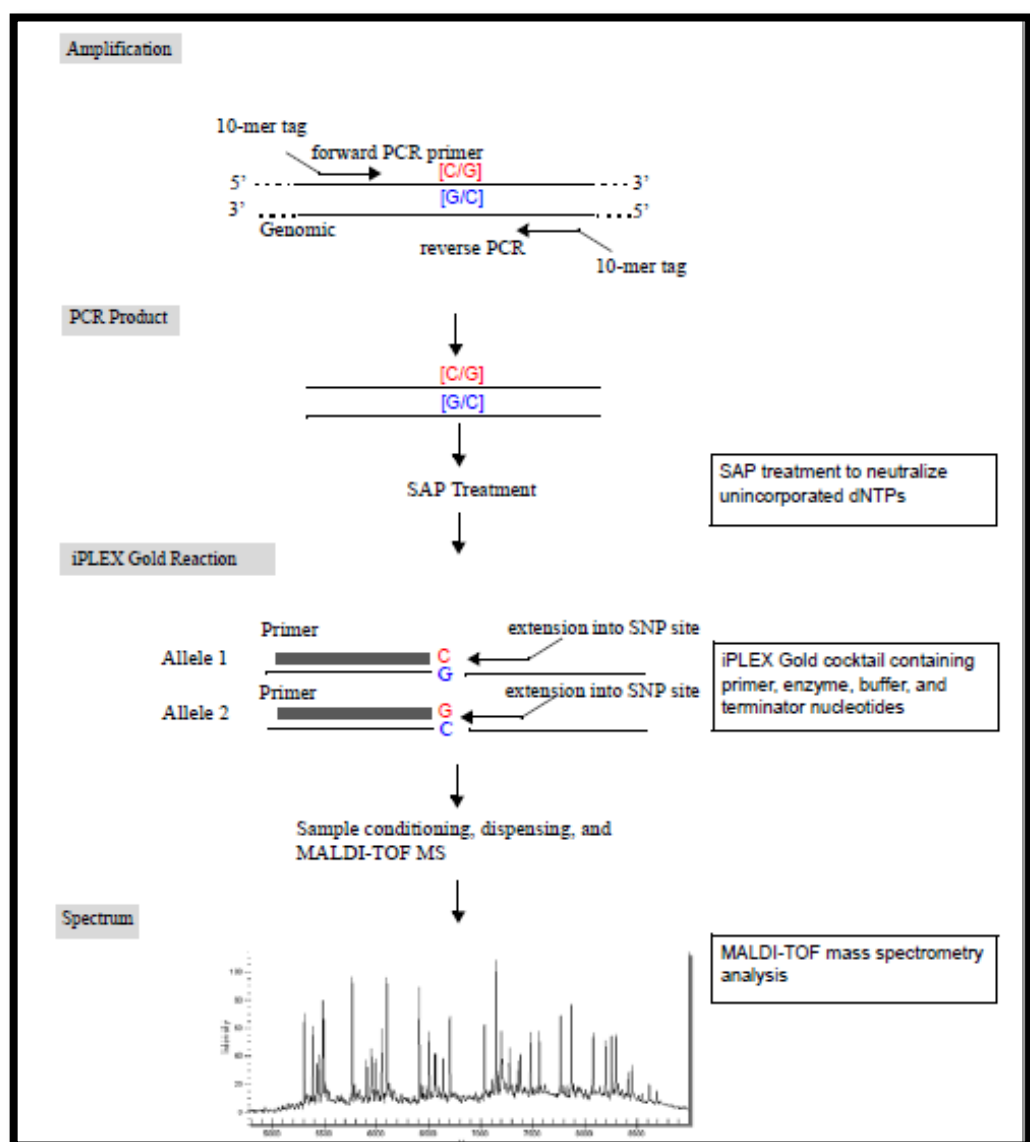
Figure 8. The iPLEX Gold assay (SEQUENOM, 2009)

3.3.2 Materials and methods

Animals

Animals used in this study were Duroc (D) and Norwegian Landrace (L). DNA samples from D (86) and L (106) were extracted by BioBank AS (Hamar, Norway).

Sequences

Sequences for this study were obtained from the patent documents:

- US/6,919,177 - PRKAG3 alleles and use of the same as genetic markers for reproductive and meat quality traits [Meat Quality, Reproduction]
- US/6,965,022 - Methods to identify swine genetically resistant to F18 E. coli associated diseases [Resistance to F18 E.coli associated diseases]
- US/7,785,778 - Porcine polymorphisms and methods for detecting them [Resistance to enterotoxigenic E.coli]
- US/2004/0126795 - Genetic markers associated with scrotal hernias in pigs [Scrotal hernias]
- WO/2007/084855 - Genetic markers for boar taint [Boar taint]

Blast search was performed for each sequence. In the case when sequence from the patent document was not of sufficient length for genotyping or was misaligned with annotated genes the correspondent sequence from GeneBank was preferred. The list of the sequences and source information are presented in Appendix 1.

Primers

Primers were designed using MassARRAY Typer software (SEQUENOM, San Diego, USA) with 4500-8000 kDa setting for molecular mass. The successful sequences were fitted into 2 multiplexes of 11 and 19 respectively, and used for further experiment. The list of primers that were used for genotyping is presented in Appendix 2.

Solutions

All solutions are presented in Appendix 3.

PCR reaction

Materials: PCR mix, genomic DNA.

Methods (for each of the multiplexes): PCR mix was prepared, lightly vortex and centrifuged. An automated liquid handling process was used for dispensing 4 µL of the PCR mix to each well of the multiplex. Genomic DNA (1 µL) was added to each well. Plates were vortexed and centrifuged at 1000 RPM.

The samples were thermocycled as follows:

```
94° C   for 15 minutes
94° C   for 20 seconds ⎤
56° C   for 30 seconds ⎬ 45 cycles
72° C   for 1 minute   ⎦
72° C   for 3 minutes
4° C    forever
```

SAP Reaction

Materials: SAP mix.

Methods (for each of the multiplexes): SAP mix was prepared, lightly vortexed and centrifuged. An automated liquid handling process was used for dispensing 2 µL of the SAP mix into each sample well. The sample plates were vortexed and centrifuged at 1000 RPM.

The sample plates were incubated as follows:

1.   37° C for 40 minutes.
2.   85° C for 5 minutes.
3.   4° C forever.

The iPLEX Gold Reaction

Materials: iPLEX Gold mix.

Methods (for each of the multiplexes): The iPLEX Gold mix was prepared, lightly vortexed and centrifuged at 1000 RPM for one minute. An automated liquid handling process was used for adding 2 µL of iPLEX Gold mix to each sample well. The sample plates were vortexed and centrifuged at 1000 RPM.

The samples were thermocycled as follows:

```
94° C   for 30 seconds
94° C   for 5 seconds  ┐
52° C   for 5 seconds  ├ 5 cycles  ┐ 40 cycles
80° C   for 5 seconds  ┘           │
72° C   for 3 minutes
4° C    forever
```

Processing of the iPLEX Gold Reaction

Materials: resin, nanopure water.

Methods (for each of the multiplexes): Resin was transferred from its container onto the dimple plate using the elongated spoon and spread into the wells of the dimple plate using the scraper. Dimple plate with the resin stood for 20 minutes. While the resin stood in the dimple plate, nanopure water (16 µL) was added to the sample plate using automated liquid handling process. After 20 minutes, the sample plate was gently placed, upside-down, onto the dimple plate, so that the resin fell out of the dimple plate and into the wells of the sample plate. The sample plate was rotated on a rotator for five minutes at room temperature and centrifuged at 3200 g for five minutes.

Dispensing onto SpectroCHIP Arrays and Analyzing Spectra

Methods (for each of the multiplexes): Nanodispensing of iPLEX Gold reaction products onto a SpectroCHIP array was performed using MassARRAY Nanodispenser. Assays and plates were set up in the MassARRAY database and spectra were acquired using the MassARRAY mass spectrometer. Spectra were analyzed using TyperAnalyzer Software.

Statistical data manipulation

Testing deviation from the HWE was performed using Pearson's chi-squared test with 1 degree of freedom. The 5 % significance level for 1 degree of freedom is 3.84, and when the χ^2 value was less than this, the null hypothesis that the population is in HWE was not rejected.

4. Results

4.1 Patents-oriented MS Access database

The result of a MS Access patent-oriented database construction is the database itself. Since the database cannot be attached to the MS Word document the decision was made to present main structural and design features of this database in the way that each person with ordinary skills will be able to construct the same database from scratch or to reconstruct existing database.

Figure 9 presents the data model that was used in the patent-oriented database construction. It is a snapshot of the standard MS Access representation of the database structure.

The design of the main form is presented in Figure 10. This form appears at the start of the database, so when user double-click at the database icon, the main-form is the first thing the user gets to.

Figure 11 shows how data from the patents document is presented for the database user and the possible ways to search for documents within internal database content.
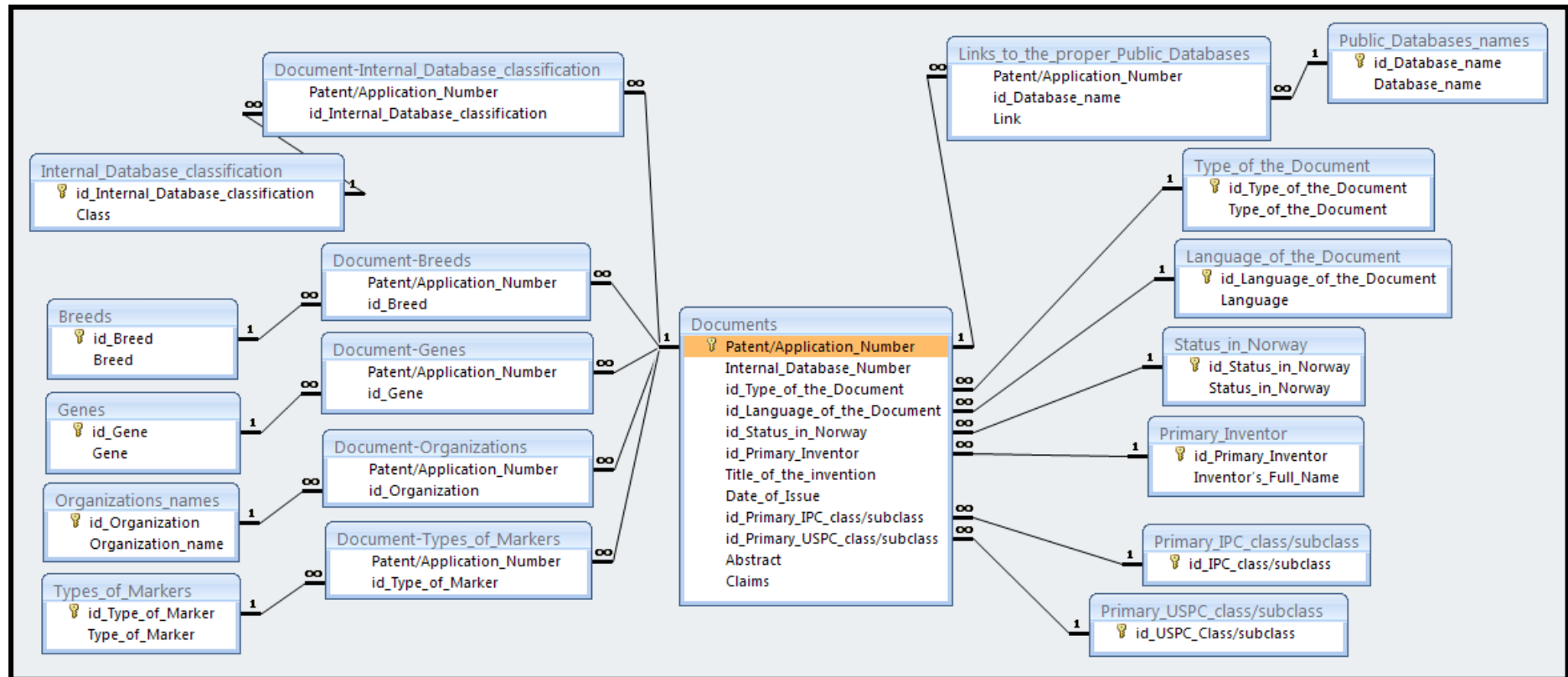
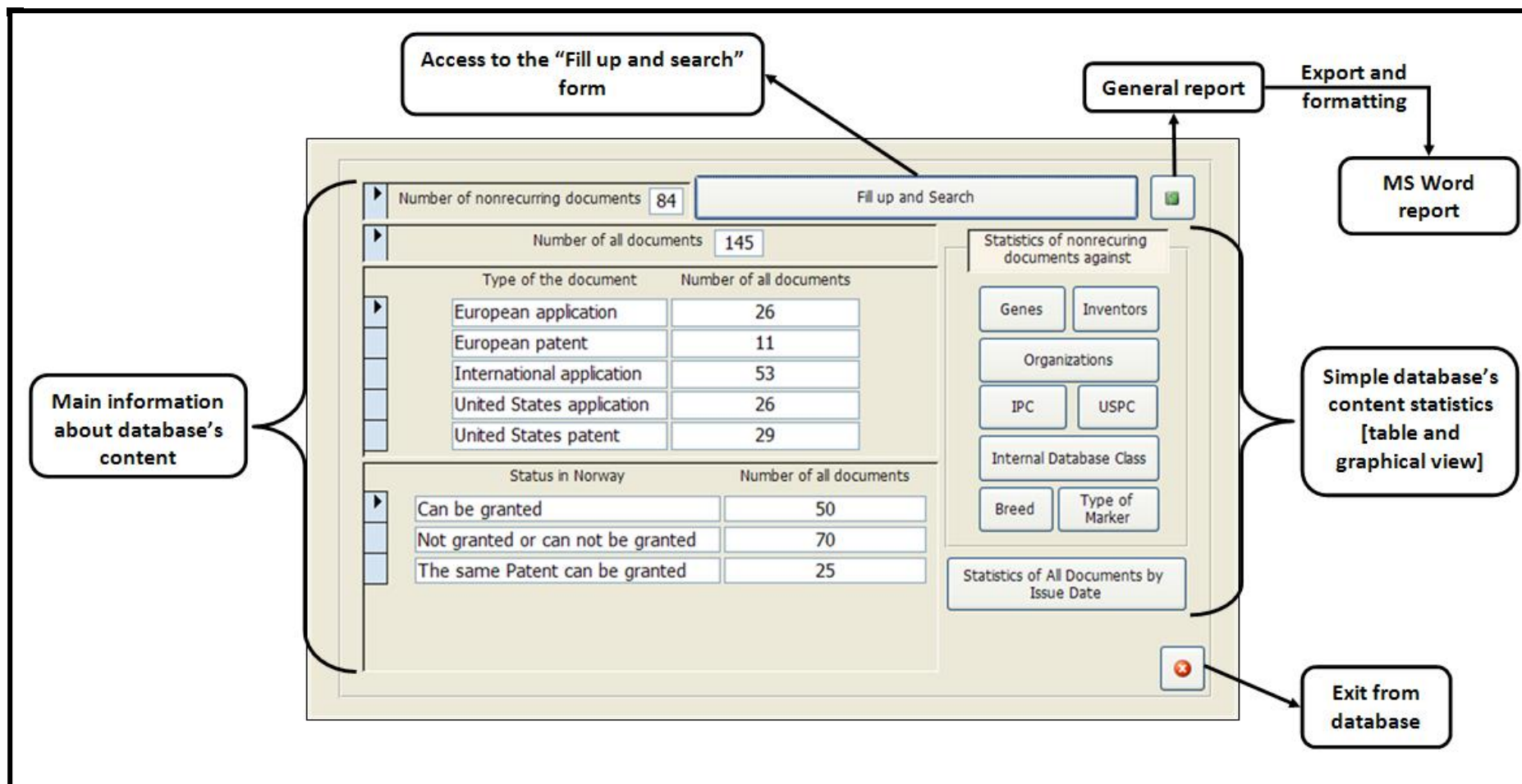*Figure 9. The data model that was used in the patent-oriented database construction*

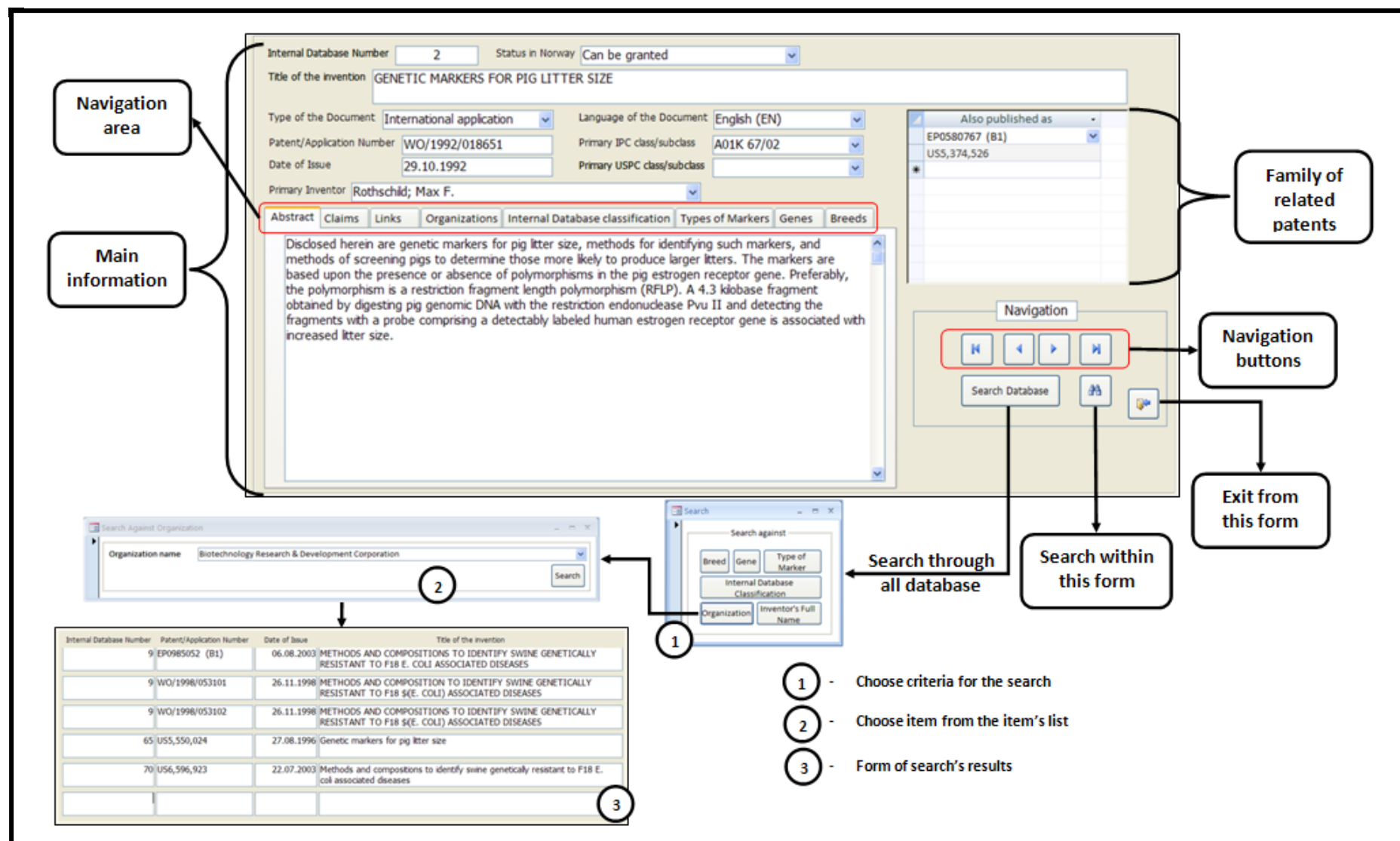*Figure 10. The design of the main form*

Figure 11. The design of the "Fill up and Search" form

4.2 Summary of patent search

All patent documents (145 documents in total and 84 nonrecurring documents) that were collected during preliminary and final patent search are shown in Appendix 4. This chapter is aimed to summarize all information about these documents

4.2.1 Statistical summary

Figure 12 is presenting simple statistics for the collected documents:

- Ranking of the documents by their IP status in Norway.
- Ranking of the documents by their origin. For those situations where group of inventors from one country filed the documents jointly with inventors or/and assignees from other countries, these countries were combined together in single origin.
- Issuing dynamic for the patent documents that were collected during this project
- Ranking of the documents by the traits that are mentioned for improving.
- Ranking of the documents by types of markers that were used in the document's experimental/example part.
- Ranking of the documents by types of breeds that were used in the document's experimental/example part.

The leader group of inventors with their assignees are listed below.

- **US + England**
  - ✓ *Inventors:* Rothschild, Max E. (Ames, IA); Tuggle, Christopher K. (Ames, IA); Bosworth, Brad T. (Littleton, NC)
  - ✓ *Organizations:* Iowa State University Research Foundation, Inc. (Ames, IA); Pig Improvement Company UK Limited (GB); The United States of America as represented by the Secretary of Agriculture (Washington, DC); Biotechnology Research & Development Corp. (Peoria, IL)
- **Canada**
  - ✓ *Inventors:* SQUIRES E JAMES (Guelph, CA)
  - ✓ *Organizations*: The University of Guelph (Guelph, CA)
- **Denmark**
  - ✓ *Inventors:* Jorgensen; Claus Bottcher (Rahavevej 1, DK)
  - ✓ *Organizations:* The University of Copenhagen (Copenhagen, DK)
- **Sweden + Belgium**
  - ✓ *Inventors:* Andersson; Leif (Uppsala, SE)
  - ✓ *Organizations:* Melica HB (Uppsala, SE); The University of Liege (Liege, BE); Seghersgentec N.V. (Buggenhout, BE)
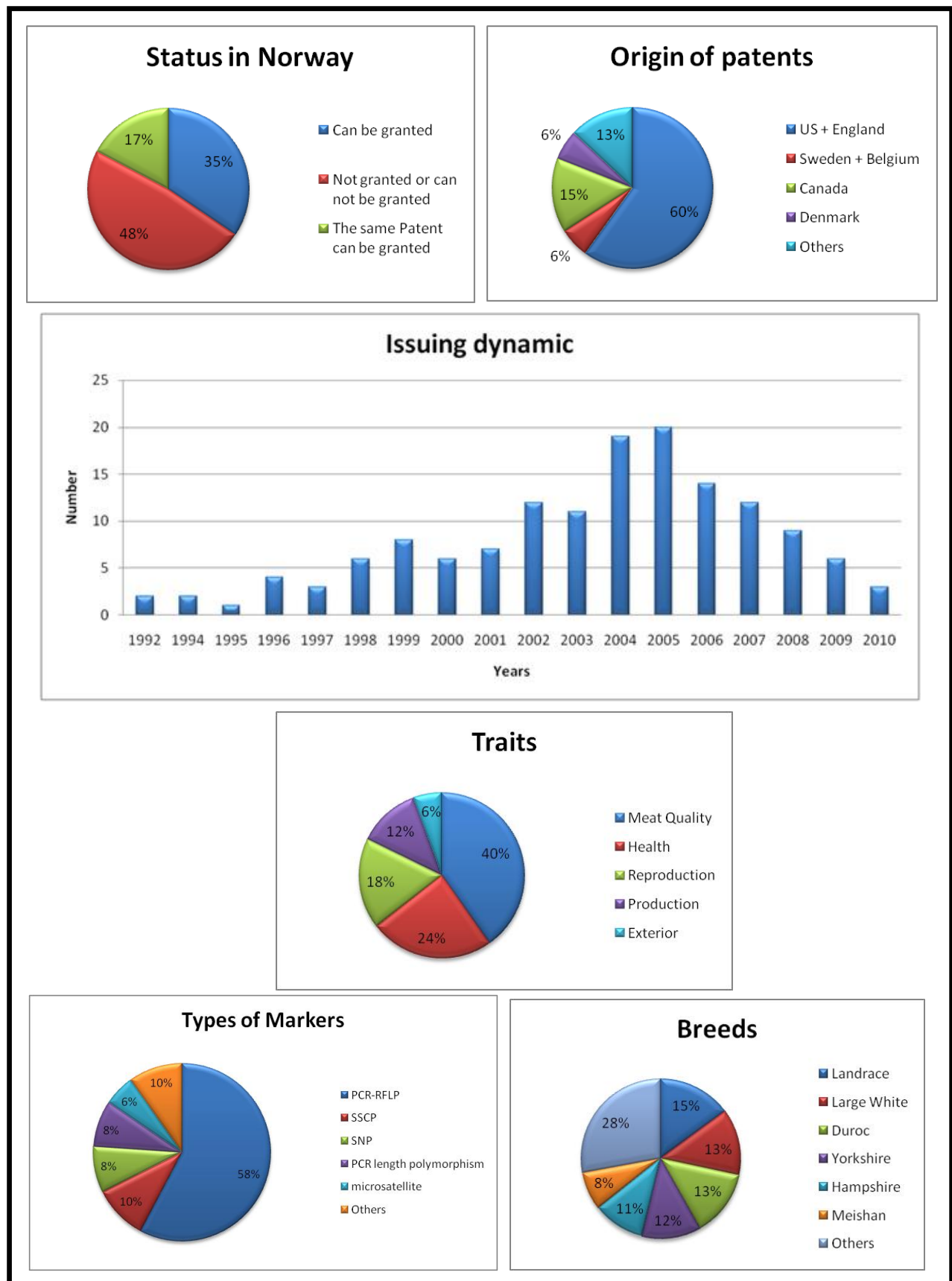
Figure 12. Simple statistics for the collected documents

4.2.2 Structure of the claims

All sets of claims from the patent documents that were collected during this search belong to the "Genes as diagnostic tools" sets of claims (Organisation for economic co-operation and development, 2002) (Figure 13). It should be noted that not all claims may be present in a single set which is dependent on the nature of the document (e.g. "test method patent", "breeding methodology patent" and so on.).
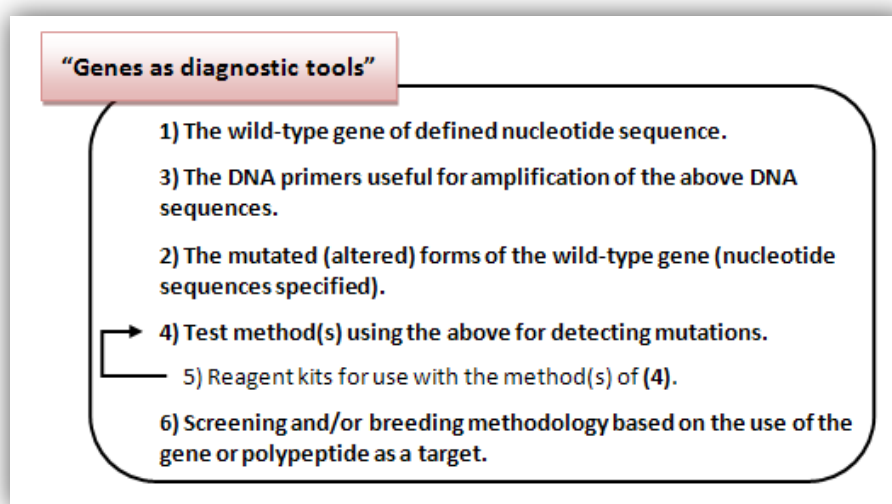
Figure 13. "Genes as diagnostic tools" sets of claims

4.2.3 Broadening claim languages

There are five general types of broadening languages that were revealed during this search, according to their linguistically structure: "translation language", "hybridization language", "percent identity language", "allelic association language" and "BLAST comparison language". It should be noted that one patent usually contain more than just one claim and inventors typically combine all these languages at once for covering as big area as possible. Short definition and example of each language are presented on the Figure 14.
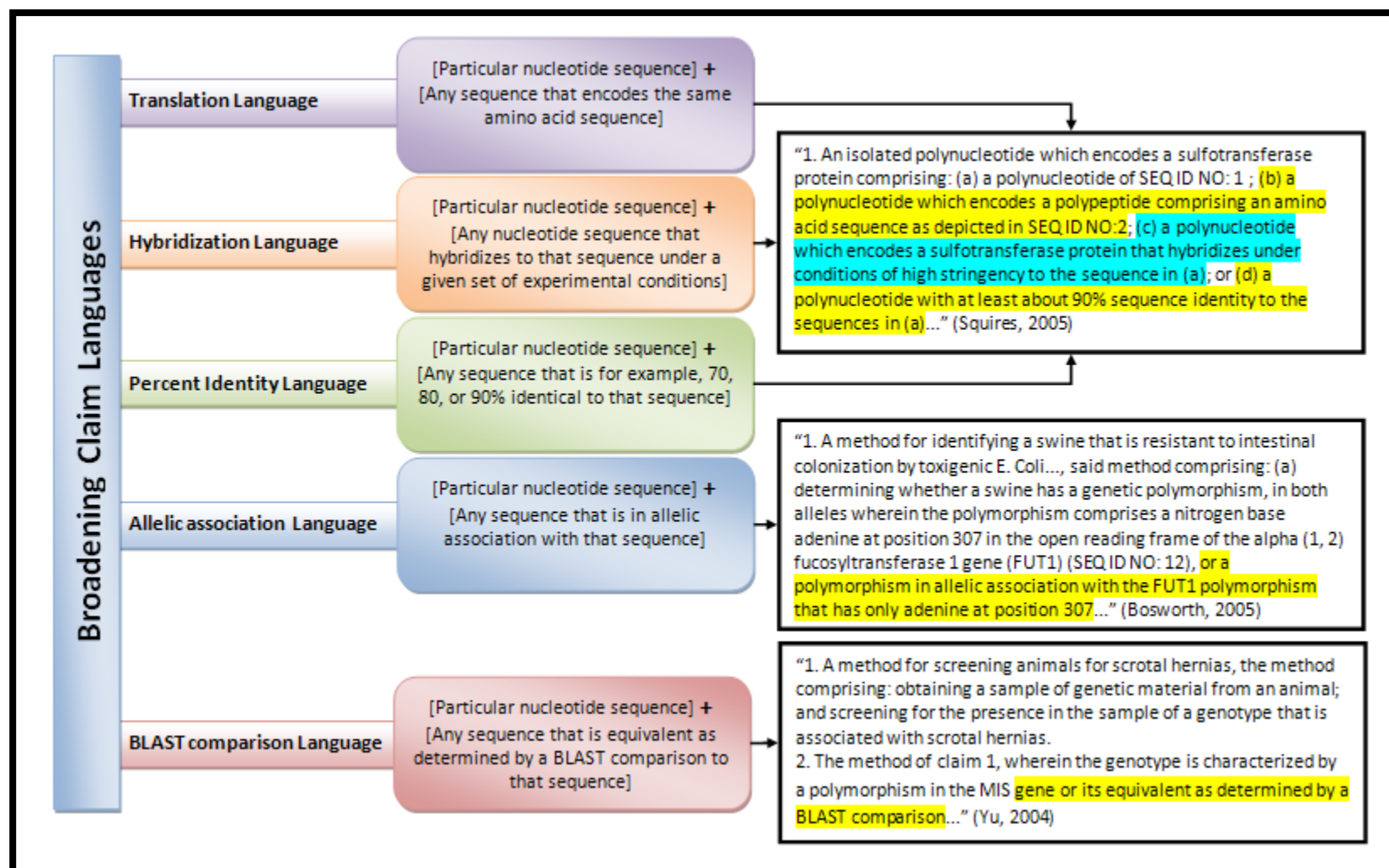
Figure 14. Broadening languages that were revealed during patent search

4.2.4 Genes coverage

Genes that were covered by the patent documents are presented in Table 5 and Table 6.

Table 5. Genes coverage

Genes		Patent №	Trait
Abbreviation	**Full Name**		
FST	Follistatin	US20080118914	Litter size
ESR	Estrogen receptor	WO/1992/018651 US5,374,526 EP0580767 (B1)	
		US6,846,632	
		US5,550,024	
PRLR	Prolactin receptor	WO/1998/003682 EP0958376 (B1)	
		US7,081,335	
		US5,935,784	
RBP4	Retinol binding protein 4	WO/2000/042218 EP1141390 (A1)	
		US5,939,264	
OPN	Osteopontin	US6,410,227 WO/1996/041892 EP0879296 (B1)	
FSHb	Follitropin subunit beta	US6,291,174	
		US20020197597	
MIS	Mullerian-inhibiting substance	US20040126795	Scrotal hernias
GPX4A	Glutathione peroxidase 4a		
RAR-gamma	retinoic acid receptor gamma	US5,939,264	Reproductive traits
MTNR1A	melatonin receptor 1a		
VCAM-*1*	vascular cell adhesion molecule 1		
PRKAG3	5'-AMP-activated protein kinase subunit gamma-3	WO/2002/020850 EP1354061 (A2)	Reproductive and meat quality
		US6,919,177	
KIT	Mast/stem cell growth factor receptor	US6,849,401 WO/1999/020795 EP1023461 (B1)	Coat colour
		US7,371,521 EP1381696 (B1) WO/2002/086158 US20040241676	
		WO/2005/108569	
		WO/1997/005278 EP0842296 (B1) US6,183,955	
HSP70.2	Heat shock protein 70	US20050112648	Backfat thickness
		US7,435,543	
MYOG	Myogenin	WO/1997/023644 US6,143,880	muscle growth
HMGA	high mobility group A family	WO/2003/078651 EP1485504 (A2)	Growth, fatness, meat quality, and feed efficiency
		US7,244,564	
H-FABP	Fatty acid-binding protein	WO/1997/035878 EP0889904 (A2)	Body weight
MC4R	melanocortin-4 receptor	US6,803,190	Fat content, weight gain, and/or feed consumption
		US20040261138 US7,303,878 WO/2001/075161 EP1276905 (A2)	
		EP1100970 (A2) WO/2000/006777	

Table 6. Genes coverage (Continue)

Genes		Patent №	Trait
Abbreviation	**Full Name**		
pLEPR	porcine leptin receptor	US6,458,531	Leanness
		US20070190527	
		WO/2005/017204 EP1651777 (B1)	
AOX1	aldehyde oxidase	WO/2005/030924 EP1623004 (A2)	Boar taint (skatole metabolism)
CYP2A6	Cytochrome P450 2A6		
CYP2E1	Cytochrome P450 2E1	WO/2007/084855 EP1984519 (A2)	
3aHSD	3 alpha hydroxysteroid dehydrogenase		
3bHSD	3 beta hydroxysteroid dehydrogenase		
SULT1A1	Sulfotransferase 1A1		
CYP17A1	Cytochrome P450 17A1		
CYTB5	Cytochrome B5	WO/1998/041861 EP0966682 (A1)	
SULT2A1	Sulfotransferase 2A1	WO/2005/074483 EP1737976 (A2)	
		US20090221539 EP1969126 (A1) WO/2007/068115	
		WO/2005/123922 EP1766025 (A1)	
		US20060024708	
P450c17	LH-stimulated 17 alpha-hydroxylase	WO/1999/018192	
NRAMP1	Natural resistance-associated macrophage protein	US20030129609	Disease resistance
		US6,844,159	
		WO/2002/097058 EP1425414 (A2)	
		US20040253594	
BPI	Bactericidal/permeability-increasing protein	US7,070,929	
		US20040234980	
		WO/2002/097136 EP1397511 (A1)	
FUT1	alpha (1,2) fucosyltransferase 1	US6,596,923	Resistant to F18 E. coli associated diseases
		EP0985052 (B1) WO/1998/053102 WO/1998/053101	
		US6,965,022	
MUC4	Myogenic factor 4	US7,785,778	Resistance to enterotoxigenic E.Coli (ETEC)
MX1	Myxovirus (influenza virus) resistance 1	US20060166188 WO/2004/038022	Resistance to RNA virus-origin disease
RYR1	ryanodine receptor	WO/1992/011387 EP0563144 (B1) US5,358,649	Malignant hyperthermia

4.3 Results of genotyping

Sequences SEQ9_[SNP_81] (MIS gene) and SEQ22_[SNP_81] (SULT1A1 gene) failed during genotyping. Results of genotyping for successful genetic markers are presented in Appendix 5 and summarized in Table 7.

Table 7. Results of genotyping

Document №	Trait	SNP_ID	Gene	Information about favourable allele	Hardy-Weinberg equilibrium	
					Landrace	Duroc
US/6,965,022	Resistance to F18 E.coli associated diseases	SEQ3_[SNP_81]	FUT1	Available	Yes	Yes
WO/2007/084855	Boar taint	SEQ11_[SNP_83]	CYP2E1	Not available	Yes	Yes
WO/2007/084855		SEQ12_[SNP_76]		Not available	Yes	Yes
WO/2007/084855		SEQ13_[SNP_81]		Not available	Yes	Yes
WO/2007/084855		SEQ14_[SNP_81]	CYP17A1	Not available	Yes	Yes
WO/2007/084855		SEQ15_[SNP_71]	CYTB5	Not available	Yes	Yes
WO/2007/084855		SEQ16_[SNP_81]		Not available	Yes	Yes
WO/2007/084855		SEQ18_[SNP_81]	3bHSD	Not available	Yes	Yes
WO/2007/084855		SEQ19_[SNP_81]		Not available	Yes	Yes
WO/2007/084855		SEQ20_[SNP_81]	SULT1A1	Not available	Yes	Yes
WO/2007/084855		SEQ24_[SNP_81]	SULT2A1	Not available	Yes	Yes
WO/2007/084855		SEQ25_[SNP_121]		Not available	Yes	Yes
WO/2007/084855		SEQ11_[SNP_82]	CYP2E1	Not available	Yes	No
WO/2007/084855		SEQ27_[SNP_78]	CYP2A6	Not available	Yes	No
US/7,785,778	Resistance to enterotoxigenic E.coli	SEQ4_[SNP_81]	MUC4	Available	Yes	No [favourable]
US/7,785,778		SEQ5_[SNP_81]		Available	Yes	No [favourable]
US/7,785,778		SEQ6_[SNP_81]		Available	Yes	No [favourable]
US/7,785,778		SEQ7_[SNP_81]		Available	Yes	No [favourable]
US/6,919,177	Meat Quality Reproduction	SEQ1_[SNP_81]	PRKAG3	Available	No [favourable]	Yes
US/6,919,177		SEQ1_[SNP_146]		Available	Homozygous [favourable]	Homozygous [favourable]
US/6,919,177		SEQ2_[SNP_87]		Available	Homozygous [favourable]	Homozygous [favourable]
WO/2007/084855	Boar taint	SEQ21_[SNP_81]	SULT1A1	Not available	Homozygous	Yes
WO/2007/084855		SEQ17_[SNP_81]	3aHSD	Not available	Homozygous	Homozygous
WO/2007/084855		SEQ23_[SNP_81]	SULT2A1	Not available	Homozygous	Homozygous
WO/2007/084855		SEQ25_[SNP_81]		Not available	Homozygous	Homozygous
WO/2007/084855		SEQ26_[SNP_81]	CYP2A6	Not available	Homozygous	Homozygous
US/2004/0126795	Scrotal hernias	SEQ8_[SNP_81]	FSHb	Available	Yes	No [unfavourable]
US/2004/0126795		SEQ10_[SNP_81]	GPX4A	Available	Yes	No [unfavourable]

5. Discussion

5.1 Last trends

Rapid progress has been made during the last twenty years in pig gene identification, mapping and functional analysis. In 1993, the public databases had only gathered approximately 600 pig sequences, while at the end of 2006, the total number of pig sequences had reached nearly 1.3 million entries (Jiang & Rothschild, 2007). In 2003, the porcine genetic linkage map had nearly 3000 loci and porcine physical genetic map contained over 3000 genes and markers (Rothschild M. F., 2003). The current annotated swine genome assembly (Sscrofa9) was released in September 2009 (Archibald, et al., 2010). By now, Pig Quantitative Trait Locus (QTL) database contains 6344 pig QTLs from 281 publications (Pig QTLdb). Some of this information is readily used by the commercial livestock companies to improve productivity and consuming properties of their populations.

The issuing dynamic that was revealed in this project allows tracing back in detail how actively the field of genetic markers in swine breeding has been covered by the patent protection during the last twenty years. From this dynamics it can be concluded that the interest to the IP of this field started from 1992.

In the 1990s, two events occurred which considerably changed the focus of the researchers in the field of swine genetics from the laboratory work to the commercialization and patenting of their inventions. First and foremost was the advent of molecular biology. New DNA based approaches gave new opportunities for exploring the genetic differences that existed within the domesticated pig. This new approach was significantly more expensive and required larger investments than previous research. Secondly, breeding companies were becoming interested in DNA based selection and were willing to fund this high-risk/high-tech research.

Perhaps the best known and largest single royalty-generating patent in animal breeding was filed like IPO application in 1992 (WO/1992/011387) and issued like US patent in 1994 (US/5,358,649). This patent claims method for screening pigs (with *HAL* 1843™ marker) to determine their susceptibility to malignant hyperthermia (MH).

The first patent that claims use of a genetic marker for a quantitative trait (pig litter size) was filed like IPO application in 1992 (WO/1992/018651) and issued like US patent in 1994 (US/5,374,526). This was also the first patent that had been exclusively assigned to one breeding company (Pig Improvement Company).

These two patents became a precedent for further patenting. Consequently issuing of patents and patent applications was increasing from 1992 until 2005. It should be noted that the impact of the second patent (US/5,374,526) is not limited just by the fact that it became possible to get a patent on the DNA based selection method. The invention that was claimed in this patent was done using "candidate gene" approach[1] and "major gene"[2] concept for selecting animals on simply inherited quantitative traits. This idea was readily utilized in the bunch of other explorations later on. In fact almost all patents and patent applications that were filed until now claim inventions that were done using "candidate gene" approach and "major gene" concept.

The greatest interest to the IP of genetic markers in swine was in years 2004-2005. A number of useful causative mutations and linked marker polymorphisms for reproduction (e.g., ESR, PRLR, RBP4, FSHb), feed intake and growth (e.g., MC4R), body composition (e.g., MYOG, H-FABP), coat colour (KIT), meat quality (e.g., HMGA, PRKAG3) and disease resistance (e.g., FUT1, BPI) were discovered and protected via patenting.

After 2005 issuing of patents and patent applications became less and less frequent. This situation can be explained as follows: Despite that the amount of information about the swine genome was increasing during 2005-2010, most of the useful major genes and markers for economically important traits related to these genes that were not kept as a trade secret had been either patented or filed like a patent application until 2007.

---

[1] "Candidate gene" approach is the approach for identification of genes that are involved in the trait's development. This approach is based on preliminary data derived from studying of model objects.

[2] "Major gene" concept is the concept that is based on the idea that some quantitative traits are mainly formed under control of a few genes (major) and other genes can be left out of account.

5.2 Present situation, future trends and their consequences

5.2.1 Present situation

The present situation is that almost all genetic markers for simply inherited qualitative characteristics governed by "major genes" (e.g. porcine stress syndrome, litter size, coat colour, resistant to specific F18 E.coli associated diseases) have been detected and patented.

More than 100 genetic markers for up to 35 genes and more than 80 methods for screening pigs with these markers were covered by the patent protection at the year 2010. All this information became the unique commercial asset for the livestock genetics companies. However, the experimental expense had limited the numbers of companies that were involved in the development of this field. There are five main companies that founded research and development in the area of swine genetic: Pig Improvement Company (PIC) UK Limited, HYPOR, TOPIG, DANBRED and Monsanto. Each of these companies had made their own decision about the way of protection for their genetic markers.

From the analysis of the patent documents it can be seen that the biggest part of all collected documents (up to 60%) were filed by scientists from the University of Iowa in the US. Almost all Iowa's patents are assigned to the PIC.

The other part of the patents is preferably owned by the universities such as The University of Guelph (Canada), University of Copenhagen (Denmark) and University of Liege (Belgium). It is interesting that big companies such as HYPOR and DANBRED probably sponsored research groups in these universities. Nevertheless they never appear on the front page of the documents and no information about licensing is available in the public sector.

PIC is the only company that made a decision to protect their intellectual assets via patenting until now and other big companies (i.e. HYPOR, TOPIG, DANBRED, and Monsanto) had as their strategy not to patent but to publish or to keep as trade secrets.

5.2.2 Future trends

The "candidate gene" approach and "major gene" concept are almost played out by now. The main future trend will most certainly be related to shifting of research interests towards low-heritable quantitative traits.

The "candidate gene" approach works with only a small part of the genome and leads to genetic tests with only 1-5 markers for effects of major genes. This number of markers was enough for selection on simply inherited qualitative traits. In contrast, exploration of low-heritable quantitative traits demands simultaneous evaluation of effects from many genes, each alone having an infinitely small contribution in the trait's development. This leads to genetic tests with more than 100 markers. Just recently it became possible to broadly use new high-throughput SNP genotyping technologies for studying low-heritable quantitative traits.

As more markers are discovered and validated by using new technology, individual marker effects become less and less important. Consequently using of this technology can potentially lead to a new field in patenting of genetic markers: patenting of selection methods which is not limited to single markers but claim broad SNP profiles. However initially patent protection was sought for individual markers and issue of patentability for multi-markers systems are yet not settled. That is why in the nearest future increasing number of markers will most probably lead to IP of inventions via trade secret instead of patenting.

There is one more possible trend that can appear and should be mentioned. In comparison to the first "deriving from the research" trend key driver for this trend came from the field of IP laws.

In the 2005, Monsanto filed one PCT applications for very extensive patents on breeding swine (WO 2005/017204). The patents are based on simple procedures, but are incredibly broad in their claims. There are more than 160 countries mentioned where the patent is supposed to be granted. WIPO already forwarded the applications to regional patent offices. At this stage the patents are not yet granted, but they could be accepted for example under European and US regulations. If these patents will be granted, it should force other companies to file the same extensive patents nature of which is very different from everything that was granted previously.

5.2.3 Potential consequences for research strategy

In General

As soon as almost all useful genetic markers were patented it was very important to figure out what potential opportunities and threats these patents can provide for the research. Of course each patent is an independent entity which scope is determined by the claims. Some patents on genetic markers may have very wide coverage across species and DNA sequences while others may be limited to single polymorphisms in one breed. Patents may be related to a process, a product produced by a process or dependent on another patent. However the overall situation can be described as follows.

Generally patents should promote research in the area of technology for which they belong. That is why patents protect but not hide technical information and some opportunities can be provided by patent study.

It should be mentioned that none of the EPO patents that was collected during this project cover Norway, which gives a great opportunity for using these patents. However the search in the Norwegian patent office was not the aim of this project which means that some details can be missed.

There are two opportunities if some similar patents exist in Norway. First opportunity is that most probably testing of these patents will fall in the term of experimental use defences. This provides a possibility for inventions around the patent. Second opportunity is that more than half of collected patents protect just methods of analysis and after precisely studying of their claims it can appear that using other methods will not infringe the patent.

Despite all opportunities that can be provided by patent study, there is one very serious threat: even if all claims' information were precisely analyzed, claims' structure and nature of the patent were determined (e.g. "method patent", "sequence patent") and all arbitrary defining terms and broadening languages were taken into account, there is a possibility that the patent holder will demand litigation procedure which will lead to the litigation costs.

From the results of laboratory testing

Conclusion about usefulness of the patented markers from the documents US/6,919,177; US/6,965,022; US/7,785,778 and US/2004/0126795 is presented in the Table 8.

Table 8. Conclusion about usefulness of the patented markers

	Resistance to E.coli						Meat Quality Reproduction			Scrotal hernias					
	FUT1			MUC4			PRKAG3			FSHb			GPX4A		
	TNA	FFA	AFA	TNA	FFA	AFA	TNA	FFA	AFA	TNA	FFA	AFA	TNA	FFA	AFA
Landrace	1	-	-	4	-	-	3	3	-	1	-	-	1	-	-
Landrace	These alleles are not fixed and can be useful for selection.						Favourable alleles are almost fixed already. These SNPs will not be very useful for selection.			These alleles are not fixed and can be useful for selection.					
Duroc	1	-	-	4	4	-	3	2	-	1	-	1	1	-	1
Duroc	These alleles are not fixed and can be useful for selection.			Favourable alleles are almost fixed already. These SNPs will not be very useful for selection.						The unfavourable allele is almost fixed in the population. This SNP will be highly useful for selection.					

TNA - Total number of SNPs.
FFA - Number of fixed favourable alleles of SNPs.
AFA - Number of fixed unfavourable alleles of SNPs.

Patent document WO/2007/084855 does not provide any information about favourable alleles for the patented markers. So nothing can be concluded in those situations when all animals were homozygote. However, most of the markers from this patent are in HWE and consequently are not fixed in Norsvin's Landrace and Duroc populations. It means that these markers can be potentially useful for selection after additional association study.

The legal status for each document is listed below:

- US/7,785,778 the similar patent application was filed in Norway (Pending) but no patents were granted until now
- US/6,965,022 the similar patent application was filed in Norway (Refused) but no patents were granted until now
- US/2004/0126795 (Assignment owner name - SYGEN INTERNATIONAL, California) no patents were granted and no patent application were filed in Norway until now
- WO/2007/084855 no patents were granted and no patent application were filed in Norway until now
- US/6,919,177 no similar patents and applications were found for Norway until now

Consequently all this patent information can be most certainly used for free (but more precise search in Norwegian patent office is still required).

5.2.4 Potential consequences for intellectual protection strategy

Obtaining and maintaining patents is a form of risk management for a company. Patents themselves do not ensure income. They must be promoted and protected. It is very important to determine when expenses associated with patent protection are justified or not. For those situations where the cost/benefit analysis indicates that the expense of a patent protection is not justified, an alternative form of risk management should be sought.

Norsvin owns no EPO or US patents and has filed no EPO or US patent applications by now. This seems logical because there is no direct competition in the Norwegian pig genetic market. Moreover, the fact that none of the EPO patents found during this project have designated Norway can indicate that for long period of time Norway was not seen as a market for big pig breeding companies.

As it was mentioned, the era of patents for many individual markers is now gone and the issue of patentability for multi-markers system are yet not settled. It means that most probably the patents authority will require dividing a multi-markers patent into several distinct patents which will destroy the nature of the patent and in general will decrease the level and increase the cost of protection for the initial (multi-markers) invention.

Obtaining a patent is a relatively expensive process. One can typically expect to spend at least $20000 per country in which protection is sought. Moreover, it is time consuming and generally takes more than 36 months for the genetic patents. It means that the genetic marker effect needs to be very large to benefit from the patent. In the case of multi-markers tests effect of each marker is relatively small, so it is obvious that expenses associated with patent protection for multi-markers tests are not justified until it will be possible to get a single patent with good determined scope.

Norsvin's present strategy of publishing instead of patenting (defensive publishing) is probably the best way of protecting information about genetic tests for now. The cost of defensive publication can be zero (e.g. conference paper) and at the same time the company can use that publication as a shield against threatened litigations and dangerous licensing campaigns.

Conclusion

Observing the trend of the last twenty years, it can be concluded that almost all patents and patent applications that were filed until now claim inventions that were done using "candidate gene" approach and "major gene" concept. This led to genetic tests with only 1-5 markers for selection on simply inherited qualitative traits.

The "candidate gene" approach and "major gene" concept are almost played out by now. The present situation is that almost all genetic markers for simply inherited qualitative characteristics governed by "major genes" have been detected and patented.

The main future trend will most certainly be related to shifting of research interests towards low-heritable quantitative traits. This leads to genetic tests with more than 100 markers, however issue of patentability for multi-markers system are yet not settled. Therefore, Norsvin's present strategy of publishing instead of patenting is probably the best way of protecting information about genetic tests for now.

None of the EPO patents that were collected during this project cover Norway, which gives a great opportunity for using these patents. However the search in the Norwegian patent office was not the aim of this project which means that some details can be missed.

As it was shown in the results of laboratory testing, despite long-time breeding history for Norsvin's L and D populations, some favourable alleles according to the claims is presented in a HWE. Moreover, 2 markers within the D population were not in HWE and unfavourable alleles according to the claims were predominant. It is indicate potential usefulness of these markers for selection.

References

Archibald, A. L., Bolund, L., Churcher, C., Fredholm, M., Groenen, M., Harlizius, B., et al. (2010 , July 19). Pig genome sequence - analysis and publication strategy.

Baylin, E. N., & Gill, J. (2005). *Effective Internet Search: E-Searching Made Easy!* Baylin Publications.

Beuzen, N. D., Stear, M. J., & Chang, K. C. (2000). Molecular markers and their use in animal breeding. *The Veterinary Journal* , 42-52.

Biotechnology Directive. (1998). Retrieved from Eur-Lex: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31998L0044:EN:NOT

Bosworth, B. T. (2005). *Patent No. US/6,965,022.*

Bryant, J. L. (1998). *Protecting Your Ideas: The Inventor's Guide to Patents.* Academic Press.

Calishain, T. (2004). *Web Search Garage.* Prentice Hall.

Cambia. (n.d.). *Broadening Claim Language in Gene Patents.* Retrieved from Patent Lens: http://www.cambia.org/daisy/RiceGenome/3660/3609.html

Crespi, S. R. (2000, Jun). Patents on genes: clarifying the issues. *Nature Biotechnology , 18* (6), pp. 683-4.

Dekkers, J. M. (2004). Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *Animal Science , 82*, 313–328.

Durham, A. L. (2004). *Patent Law Essentials: A Concise Guide Second Edition.* Praeger Press.

Engelfriet, A. (2005, October 5). *Determining the scope of a patent.* Retrieved from Ius mentis: http://www.iusmentis.com/patents/claims/

Engelfriet, A. (2005, October 1). *Differences between US and European patents.* Retrieved from Ius mentis: http://www.iusmentis.com/patents/uspto-epodiff/

Engelfriet, A. (2002, March 12). *Frequently Asked Questions: General questions.* Retrieved from Ius mentis: http://www.iusmentis.com/patents/faq/general/

European Patent Convention. (1973). Retrieved from European Patent Office: http://www.epo.org/law-practice/legal-texts/html/epc/1973/e/ma1.html

Fuller, L. U. (2007). *Access 2007 for Dummies.* John Wiley & Sons, Inc.

Groh, M. R. (2007). *Access 2007 Bible.* John Wiley & Sons, Inc.

Gura, S. (2007). Livestock Genetics Companies.

Hawkins, N. (2010, December). Human Gene Patents and Genetic Testing in Europe: A Reappraisal. *SCRIPTed , 7* (3), pp. 453-473.

Hunt, D., Nguyen, L., & Rodgers, M. (2007). *Patent Searching: Tools & Techniques.* Landon.

Jauhar, A., & Narnaulia, S. (2010, January). Patenting Life the American, European and Indian Way. *Journal of Intellectual Property Rights , 15*, pp. 55-65.

Jennings, R. (2007). *Using Microsoft Access 2007.* Que.

Jiang, Z., & Rothschild, M. F. (2007). Swine Genome Science Comes of Age. *International Journal of Biologicak Science , 3*, pp. 129-131.

John, J. D. (1998, May 1). The Patenting of DNA. *Science , 280* (5364), pp. 689-690.

Lakshmikumaran, M. (2007, January). Patenting of Genetic Inventions. *Journal of Intellectual Property Rights , 12*, pp. 45-56.

Lambe, N., & Simm, G. (2004). ANIMAL BREEDING AND GENETICS/Traditional Animal Breeding. *Encyclopedia of Meat Sciences* , 11-19.

Lambert, S., & Lambert, M. D. (2007). *Microsoft Office Access 2007 Step by Step.* Microsoft Press.

McLaren, D. G. (2007). Recent Developments in Genetic Improvement of Pigs. *Manitoba Swine Seminar* .

Miller, M. (1999). *The complete idiot's guide to online search secrets.* Indianapolis, Ind.

Navajas, E. A., & Simm, G. (2004). ANIMAL BREEDING AND GENETICS/DNA Markers and Marker-Assisted Selection. *Encyclopedia of Meat Sciences* , 19-27.

Organisation for economic co-operation and development. (2002). *Genetic inventions, intellectual property rights and licensing practices.* OECD.

Patent Cooperation Treaty. (1970). Retrieved from World Intellectual Property Organization: http://www.wipo.int/pct/en/texts/articles/atoc.htm

Pig QTLdb. (n.d.). Retrieved from Pig QTLdb: http://www.animalgenome.org/cgi-bin/QTLdb/SS/index

Rothschild, M. F. (2003, April). Advances in pig genomics and functional gene discovery. *Comparative and Functional Genomics , 4*, pp. 266–270.

Rothschild, M., & Newman, S. (2002). *Intellectual Property Rights in Animal Breeding and Genetics.* CAB International.

Rothschild, M., Stalde, K., & Dekkers, J. (2010, April 15). *Genomic Selection Shows Promise in Swine Breeding.* Retrieved from National Hog Farmer:
http://nationalhogfarmer.com/genetics-reproduction/evaluation-selection/genomic-selection-tools-0426/

SEQUENOM. (2009). *iPLEX Gold Application Guide.*

Squires, J. E. (2005). *Patent No. WO/2005/123922.*

Strasbourg Agreement. (1971). Retrieved from WIPO:
http://www.wipo.int/wipolex/en/wipo_treaties/details.jsp?treaty_id=11

TRIPS Agreement. (1994). Retrieved from World Trade Organization:
http://www.wto.org/english/tratop_e/trips_e/t_agm0_e.htm

U.S. Code Title 35. (n.d.). *United States Code.* Retrieved from United States Government Printing Office:
http://frwebgate.access.gpo.gov/cgi-bin/usc.cgi?ACTION=BROWSE&title=35usc

U.S. Department of Commerce. (2008). *Patents and How to Get One: A Practical Handbook.* Dover.

Yu, T.-P. (2004). *Patent No. US/2004/0126795.*