

UNIVERSIDADE ESTADUAL DE SANTA CRUZ
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E
BIOLOGIA MOLECULAR



SISTEMA DE INFORMAÇÃO PARA APOIO A PROGRAMAS DE
MELHORAMENTO GENÉTICO BASEADOS EM MARCADORES
MOLECULARES: APLICAÇÃO AO CASO DO CACAUEIRO
(*Theobroma cacao* L.)

LÍVIA ARAÚJO LESSA SERTÓRIO DE SOUZA

ILHÉUS – BAHIA – BRASIL

Junho de 2005

LÍVIA ARAÚJO LESSA SERTÓRIO DE SOUZA

**SISTEMA DE INFORMAÇÃO PARA APOIO A PROGRAMAS DE
MELHORAMENTO GENÉTICO BASEADOS EM MARCADORES MOLECULARES:
APLICAÇÃO AO CASO DO CACAUEIRO (*Theobroma cacao* L.)**

Dissertação apresentada à
Universidade Estadual de Santa Cruz,
como parte das exigências para
obtenção do título de Mestre em
Genética e Biologia Molecular.

LÍVIA ARAÚJO LESSA SERTÓRIO DE SOUZA

ILHÉUS – BAHIA – BRASIL

Junho de 2005

S729

Souza, Lívia Araújo Lessa Sertório de.

Sistema de informação para apoio a programas de melhoramento genético baseados em marcadores moleculares : aplicação ao caso do cacau (Theobroma cacao L.) / Lívia Araújo Lessa Sertório de Souza. – Ilhéus, Ba : UESC, 2004.

95f. : il.

Orientador: Diego Gervasio Frias Suarez

Dissertação (Mestrado) – Universidade Estadual de Santa Cruz. Programa de Pós-graduação em Genética e Biologia Molecular

Inclui bibliografia e apêndices.

1. Cacau – Melhoramento genético. 2. Sistemas de informação. 3. Marcadores biológicos. 4. Sistemas de recuperação da informação. I. Título.

CDD 633.74

LÍVIA ARAÚJO LESSA SERTÓRIO DE SOUZA

SISTEMA DE INFORMAÇÃO PARA APOIO A PROGRAMAS DE MELHORAMENTO
GENÉTICO BASEADOS EM MARCADORES MOLECULARES: APLICAÇÃO AO
CASO DO CACAUEIRO (*Theobroma cacao* L.)

Dissertação apresentada à
Universidade Estadual de Santa Cruz,
como parte das exigências para obtenção
do título de Mestre em Genética e Biologia
Molecular.

APROVADA: 27 de setembro de 2004

Prof^a. Dr^a. Vânia Cordeiro da Silva
(UESC)

Dr. Uilson Vanderlei Lopes
(CEPEC/CEPLAC)

Prof. Dr. Diego Gervasio Frías Suarez
(UESC – Orientador)

DEDICATÓRIA

Dedico este trabalho a meus queridos pais, Tobias e Norma, minha amada irmã, Leila e meu estimado esposo, Vinícius, que muito me incentivaram, dispensando carinho, cuidados e atenção, e compreendendo minha ausência; aos meus colegas de curso, motivo maior de minha perseverança e alegria pela convivência; e aos amigos e inimigos, forças motrizes para as ações de minha vida.

AGRADECIMENTOS

Agradeço a Deus pela vida e, com ela, a oportunidade de progredir.

À FAPESB pelo apoio financeiro a este projeto.

À CAPES pela concessão de bolsa de mestrado.

À UESC, através do Colegiado do Programa de Pós-Graduação e Genética e Biologia Molecular, pela oportunidade da realização do curso.

Ao meu Orientador, o Prof^o Dr. Diego Gervásio Frias Suarez, pelos valiosos ensinamentos e pelo apoio.

Ao Prof^o Dr. Ronan Xavier Corrêa, pela valiosa co-orientação, pelo zelo com o trabalho e pela intensiva e valiosa colaboração.

À Prof^a Dr^a. Mônica Bertão pela dedicação intensiva, quando à frente da coordenação do colegiado do curso, pela gratificante convivência e pelos valiosos conselhos.

Ao Prof^o Dr. Marco Antônio, pelo empenho em dar seguimento aos trabalhos na coordenação do curso e pelo incentivo e apoio.

Ao Corpo Docente do Mestrado pelos ensinamentos e convívio, em especial, ao Prof^o Dr. Leandro Lopes Loguércio, pelo incentivo, pelo convívio e pela amizade.

Aos pesquisadores que auxiliaram neste trabalho, em particular, os Drs. Dário Ahnert, Dr. Uilson Lopes e Dr. Milton Macoto, pelas preciosas sugestões.

Aos funcionários e alunos estagiários do Laboratório de Bioinformática, especialmente, a Ramon Vidal, a Josenaldo e a Fernanda Almeida, pelas sugestões, pelo incentivo e pela amizade.

A Tiago Motta, pelo auxílio intensivo, pelas sugestões e recomendações valiosíssimas, pela paciência e compreensão por minha insipiência na área de estudo e pela grata amizade.

A Rafael Botelho, brilhante companheiro de trabalho, pela dedicação, pela imensa colaboração, auxiliando no desenvolvimento e na correção deste trabalho, pelas maravilhosas sugestões, pelo incentivo constante, pela cumplicidade e pela amizade.

Às funcionárias do Colegiado do Curso, Andréa e Geysa, e da Pró-Reitoria de Pós-Graduação, Alzira Quadros e Cláudia, pelo saudável e alegre convívio.

A Adriana, Alfredo, Bianca, Cristiane, Geruza, Jeiza, Joci, Patrícia, Ronaldo e Valéria, pelo coleguismo, pela cumplicidade, pelo incentivo persistente, pela gratificante convivência e pela amizade.

"Embora ninguém possa voltar atrás e fazer um novo começo, qualquer um pode começar agora e fazer um novo fim".

Chico Xavier

"Os computadores são incrivelmente rápidos, precisos e burros; os homens são incrivelmente lentos, imprecisos e brilhantes; juntos, seu poder ultrapassa os limites da imaginação".

Albert Einstein

LISTA DE FIGURAS E TABELAS

Classificação de SI conforme o propósito	10
Funcionamentos de Sistemas de informação baseado na WEB	12
Componentes que formam a estrutura da arquitetura de um Sistema de Informação baseado na WEB	13
Relacionamento entre os componentes e as camadas que formam a arquitetura de sistemas de informação	16
Visão gerencial e visão técnica de desenvolvimento de software.....	18
Etapas de um Ciclo de Modelo Espiral	23
Modelo Iterativo Incremental.....	24
Diagrama de Contexto de Aplicação do Sistema de Informação	40
Casos de Uso e suas respectivas funcionalidades	41
Diagrama de Casos de Uso	42
Página principal da Interface.....	44
Formulário de interface para manipulação de informações de acessos de cacaqueiro	44
Diagrama Entidade-Relacionamento	45
Diagrama de Classes.....	47
Diagrama de Componentes	48
Projeto físico de do banco de dados	49
Representação de códigos para a construção do BD utilizando SGBD MySQL	50
Comando para a inserção de dados no BD	51
Comando para consultar informação do BD	51

LISTA DE ABREVIATURAS E SIGLAS

AFLP – Amplified Fragment Length Polymorphism

BD – Banco de Dados

BDr – Banco de Dados Relacional

HTML – Hypertext Markup Language

PHP – Hypertext Preprocessor

RAPD – Random Amplified Polymorphic DNA

RFLP – Restriction Fragment Length Polymorphism

SAS – SAS INSTITUTE INC. Software estatístico.

SCAR – Sequence Characterized Amplified Regions

SGBD – Sistema Gerenciador de Banco de Dados

SGBDr – Sistema Gerenciador de Banco de Dados Relacional

SI – Sistema de Informação

SO – Sistema Operacional

SQL - Structured Query Language

SSR – Simple Sequence Repeats (microsatélites)

TI – Tecnologia de Informação

URL – Uniform Resource Locator

VNTR – Variable Number of Tandem Repeats (minissatélites)

ÍNDICE

Extrato	xi
Abstract	xiii
1 INTRODUÇÃO	1
2 REVISÃO DE LITERATURA	4
2.1 Conceitos Básicos de Informática	4
2.1.1 Definição	4
2.1.2 Hardware	5
2.1.3 Software	5
2.1.3.1 Softwares Básicos	5
2.1.3.2 Softwares Aplicativos	6
2.1.4 Linguagens	7
2.1.5 Sistemas de Informação	7
2.1.5.1 Características e Finalidades de Sistemas de Informação	9
2.1.5.2 Sistema de Informação baseado na WEB	10
2.2 Arquitetura e Desenvolvimento de Software	11
2.2.1 Arquitetura dos Sistemas de Informação	11
2.2.2 Processo de Desenvolvimento de Software Orientado a Objeto	15
2.2.2.1 Modelos de Ciclos de Desenvolvimento	20
2.2.2.1.1 Modelo Cascata	20
2.2.2.1.2 Modelo de Desenvolvimento Evolucionário	20
2.2.2.1.3 Modelos Iterativos	21
2.3 Caracterização dos Programas de Melhoramento Genético Baseados em Marcadores Moleculares	23

2.3.1	Tipos de Marcadores Moleculares	27
2.3.1.1	RFLP	28
2.3.1.2	Minissatélites.....	28
2.3.1.3	RAPD	28
2.3.1.4	SCAR	29
2.3.1.5	AFLP	29
2.3.1.6	Microssatélites.....	29
2.4	Descrição do Domínio das Informações	30
2.5	Bancos de Dados de Marcadores Moleculares	31
3	METODOLOGIA	33
3.1	MATERIAIS	33
3.1.1	Hardware	33
3.1.2	Softwares e Linguagens	33
3.1.2.1	GNU Linux.....	33
3.1.2.2	UML.....	34
3.1.2.3	Poseidon	34
3.1.2.4	MySQL	34
3.1.2.5	PHP.....	34
3.1.2.6	JavaScript	35
3.1.2.7	Fireworks.....	35
3.1.2.8	Corel Suite	35
3.1.2.9	Dreamweaver.....	35
3.1.2.10	Programa GENES.....	36
3.2	MÉTODOS	36
3.2.1	Modelo de Desenvolvimento.....	36
3.2.2	Levantamento de Requisitos	37
3.2.3	Análise	38
3.2.4	Projeto	38
3.2.5	Implementação	39
3.2.6	Testes	39
4	RESULTADOS E DISCUSSÃO	40
4.1	Levantamento de Requisitos	40
4.1.1	Diagrama de Contexto	41

4.1.2	Modelo de Casos de Uso.....	42
4.1.3	Glossário.....	44
4.1.4	Protótipo da Interface.....	45
4.2	Análise.....	47
4.2.1	Diagrama de Entidades e Relacionamentos (DER).....	47
4.2.2	Diagrama de Classes.....	48
4.3	Projetos	49
4.3.1	Diagrama de Componentes	49
4.3.2	Projeto de Banco de Dados	49
4.4	Implementação	51
4.4.1	Banco de Dados MySQL	51
4.4.2	Manipulação de dados do Banco de Dados	51
4.5	Testes.....	52
5	CONCLUSÕES.....	54
6	REFERÊNCIAS BIBLIOGRÁFICAS	57
	GLOSSÁRIO	61
	APÊNDICES	68
	APÊNDICE 1 – Casos de Uso	69
	APÊNDICE 2 – Projeto de Banco de Dados	76
	APÊNDICE 3 – Código gerador do Banco de Dados MySQL.....	80
	APÊNDICE 4 – Formulário Semi-estruturado	85
	APÊNDICE 5 – Lista de Acessos de Cacau	95

EXTRATO

LESSA, Livia A. M.S., Universidade Estadual de Santa Cruz, Ilhéus, setembro de 2004. **Sistema de Informação para Apoio a Programas de Melhoramento Genético Baseado em Marcadores Moleculares: Aplicação ao Caso do Cacaueiro (*Theobroma cacao*)**. Orientador: Diego Gervásio Frías Suarez. Co-orientador: Ronan Xavier Corrêa. Colaborador: Júlio C. M. Cascardo.

Sistemas de Informação (SI) são conjuntos de ferramentas computacionais projetadas para processamento de informações. Diversos pesquisadores envolvidos em programas de melhoramento genético reclamam da falta de informações relevantes na área, devidamente classificadas e organizadas. Para encurtar a distância entre a geração de informações e sua utilização prática e funcional, o objetivo deste trabalho foi desenhar um SI baseado na WEB, capaz de armazenar e processar dados fenotípicos/agronômicos, genotípicos e bibliográficos, provenientes dos estudos com marcadores moleculares. Além do projeto desse sistema, que terá interface WEB disponibilizada no portal do Laboratório de Bioinformática (LABBI/UESC), foi proposta a construção de um banco de dados (BD) relacional para armazenar dados sobre germoplasma de cacau. Utilizou-se a linguagem padrão UML (Unified Modeling Language) para a modelagem dos dados, o projeto do sistema, a especificação de requisitos e para a documentação do sistema. Para o desenvolvimento do mesmo, adotou-se um modelo iterativo de tipo incremental por camadas, cuja seqüência padrão de atividades (*workflows*) de cada camada consiste em: levantamento de requisitos, análise, projeto, implementação, testes e liberação. A partir do levantamento de requisitos, foram criados o diagrama de contexto de aplicação do SI, o modelo de Casos de Uso, baseado na especificação formal de requisitos, e o protótipo da interface. A análise resultou na criação do Diagrama

Entidade-Relacionamento, o qual exprimiu o domínio da informação a ser manipulada pelo sistema e o Diagrama de Classes, que permitiu representar as entidades identificadas no modelo conceitual e as relações existentes entre elas. No projeto foi diagramada a estrutura do sistema em nível de componentes e esquematizado o projeto físico do BD. O BD relacional foi implementado utilizando-se MySQL, e a verificação de funcionalidade e integridade referencial está em fase de teste α . O uso de dados reais confere credibilidade aos testes do sistema, permitindo a avaliação do cumprimento dos requisitos relacionados com o armazenamento das informações. O principal objetivo deste trabalho foi atingido, pois na primeira iteração foi realizada a construção de um BD relacional com informações de cacau e a segunda iteração, em andamento, está abordando a camada de interface, a qual tem como característica marcante a diferenciação em níveis de abordagem, levando em conta que diferentes usuários (pesquisadores e produtores) têm necessidades distintas de informação, para auxiliar na tomada de decisão. A camada de lógica de aplicação será objeto da próxima iteração e realizada com o uso da linguagem PHP. O sistema gerado é de extrema relevância e muito proveitoso para instituições de pesquisa e agricultores, pois é um sistema pioneiro no Brasil para marcadores moleculares e propiciará a colaboração entre pesquisadores e produtores disseminando, através da internet, os resultados das pesquisas, tornando-se coadjuvante no aperfeiçoamento dos programas de melhoramento genético.

Palavras-chave: banco de dados, portal WEB, melhoramento genético, cacau.

ABSTRACT

LESSA, Livia A. Universidade Estadual de Santa Cruz, Ilhéus, September 2004. **Information System for Support to Genetic Breeding Program Based in Molecular Markers: Application to Cocoa's Case (*Theobroma cacao*)**. Advisor: Diego Gervásio Frias Suarez. Advisor Committee Members: Ronan Xavier Corrêa and Júlio C. M. Cascardo.

Information Systems (IS) are a conjunct of computation tools projected to process information. Many researchers involved in genetic breeding programs complain because exist little important information in the area properly classified and organized. To shorten the distance between the information's generation and its practical and functional utilization, the objective of this work was to draw a IS based in the WEB, capable of storing and processing, phenotypic, genotypic and bibliographic data, coming from molecular markers studies. Besides the project of this system, that will have a WEB interface and will be available in the homepage of the Bioinformatic Laboratory (LABBI/UESC), it was proposed the construction of a relational database (DB) to store cocoa germoplasm data. It was made use of the pattern language UML (Unified Modeling Language) to model the data, the system's project, the specification of requirement and the system's documentation. For the development of this system, an iterative model of incremental type by layers, which pattern sequence of activity (*workflows*) of each layer consists in: surveying the requirements, analysis, project, implementation, testes and release. From the requirements, it was created a diagram of context of the application of IS, the Case Use model, based in the formal specification of requirement, and the interface prototype. The analysis resulted in the creation of the Diagram Entity -Relationship, which expressed the information domain to be manipulated by the system and the

Diagram of Classes, which allowed to represent the entity identified in the conceptual model and the relations existent between them. In the project it was create the diagram of the system structure at the level of components and made the scheme of the physical project of the DB. The relational DB was implemented using MySQL, and the verification of functionality and referential integrity in α - test stage. The use of real data confers credibility to system tests, allowing the evaluation of the accomplishment of the related requirement with the information storage. The principal objective of this job was reached, since that in the first iteration it was constructed a relational DB with cocoa information and the second reiteration, in course, is making an approach to the interface layer, which has as key characteristic the differentiation at the level of information, taking into account that different users (researchers and producers) have different necessities of information, to help out in the decision capture. The logic layer of application will be object of the next iteration and will be done using the PHP language. The system created is of extreme importance and very advantageous for researches institution and agriculturists, therefore it is a pioneer molecular markers system in Brazil and it will allow the collaboration between researchers and producers disseminating, through the internet, results of research, contributing for in the improvement of the genetic breeding programs.

Keywords: database, portal WEB, genetic breeding, cocoa.

1. INTRODUÇÃO

O Século XX foi o período em que as tecnologias mais extraordinárias tornaram-se realidade. A explosão do desenvolvimento científico afetou profundamente a cultura e a organização social a ponto de se poder identificar uma nova era: a era da Tecnologia. O homem da era moderna vivenciou o advento do telefone e do automóvel, a construção e o uso da bomba atômica e da energia nuclear e, mais recentemente, o desenvolvimento acelerado da informática e da biotecnologia. Um mundo mediado por artefatos tecnológicos, caracterizados principalmente pelo domínio da Física Atômica e Nuclear, da Engenharia Genética e da Tecnologia da Informação, os quais formam nossas relações e definem nossas eleições e estilos de vida.

Com a descoberta da estrutura do DNA em 1953, nascem a biotecnologia moderna e a biologia molecular, agregando um conjunto de tecnologias que tem auxiliado, entre outros, os programas de melhoramento genético ou programas para preservação de plantas e animais de interesse econômico (GRIFFITHS, 2001).

O melhoramento genético é um ramo da pesquisa agrícola, médica e até ambiental que, partindo das leis de Mendel, aplicando a estatística e o conhecimento de biologia, bioquímica, biofísica e bioinformática¹, procura alterar indivíduos de uma população para que passem a expressar características desejadas (SILVEIRA, 2003). Assim, os programas de melhoramento genético constam de três fases cíclicas: (1) produção de variabilidade

¹ Toda palavra ou termo que apresenta uma pequena bola sobrescrita, possui explicação no glossário.

genética* a partir de indivíduos selecionados da população em estudo, (2) análise da diversidade genética resultante, e (3) seleção dos indivíduos mais adaptados às mudanças ambientais, mais eficientes na utilização de nutrientes, mais tolerantes a pragas e doenças e/ou mais produtivos e de melhor qualidade, segundo as finalidades do programa (LEMOS, 2003).

O uso de marcadores moleculares* para caracterizar genotipicamente os indivíduos das populações em estudo traz uma nova dimensão aos programas de melhoramento genético. Com o crescente uso desde tipo de técnica nesses programas faz-se necessário a implementação de ferramentas computacionais e procedimentos para a aquisição e o armazenamento apropriado dos novos dados.

Se por um lado o avanço tecnológico coloca à disposição uma verdadeira avalanche de informações, por outro se deve pensar em como filtrá-las, armazená-las e torná-las disponíveis. É neste contexto que vemos informações valiosas misturadas com outras irrelevantes e nos defrontamos com a necessidade de distingui-las, interpretá-las e justificá-las. Na verdade, não tem sentido acumular informações se as mesmas não forem úteis para a tomada de decisões.

As ferramentas digitais projetadas para processamento de informações

(entrada, manipulação, armazenamento, saída) em forma extensiva são chamadas de sistemas de informação. Os Sistemas de Informação (SI) são formados por uma série de elementos ou componentes de hardware e software inter-relacionados estrutural e funcionalmente, que executam as operações necessárias para coletar e disponibilizar as informações de interesse (STAIR, 1998).

Freqüentemente, os pesquisadores atuando em melhoramento genético armazenam seus dados em arquivos de texto, formatados para uso por programas estatísticos ou de processamento específico, sendo comum também o armazenamento em planilhas de cálculo. Dado que nenhuma das alternativas anteriores é apropriada para o armazenamento e a manipulação eficiente e confiável de dados, com freqüência surgem os seguintes problemas: (1) um mesmo dado é depositado repetidamente em diferentes locais, podendo provocar inconsistências se somente uma das cópias do dado ser alterada, e ocupa espaço de armazenamento desnecessário; (2) os dados coletados ou gerados por um pesquisador utilizando formato, codificação e denominação própria, são de difícil consulta por outros pesquisadores complicando também a divulgação e uso posterior dos resultados; (3) o tempo requerido para gerar os arquivos de dados de entrada de outros softwares, geralmente

em formatos e configurações específicas e diferentes, ou mesmo para sua inclusão em bancos de dados estruturados é muito elevado; e (4) a combinação de dados de muitos experimentos para uma análise conjunta é muito difícil.

Para evitar esses inconvenientes, foi proposto como objetivo principal do presente trabalho o desenho de um Sistema de Informação genérico, baseado num banco de dados relacional acessível via WEB, que permita vincular os dados genotípicos e fenotípicos das diversas variedades consideradas nos programas de melhoramento genético de plantas, assim como associá-los às características agrônômicas e referenciá-los às publicações científicas correspondentes.

Neste contexto, faz-se relevante a escolha de uma estrutura de dados⁴ apropriada para o armazenamento das informações essenciais para essa atividade incluindo-se informações de marcadores moleculares.

Segundo os conhecimentos da autora, existem poucos trabalhos publicados analisando o problema do armazenamento de marcadores moleculares para uso em programas de melhoramento de plantas, como o trabalho de FRISCH, LAMKEY e MELCHINGER (2002). Neste artigo os autores propõem uma estrutura de dados genérica para plantas, de um modo geral, construída sob paradigma relacional cuja entidade básica é a observação de presença ou ausência de alelo⁵ num loco marcador para uma amostra de DNA em estudo.

Contudo, como a proposta de FRISCH, LAMKEY e MELCHINGER (2002) trata do armazenamento apenas de dados moleculares, neste trabalho foi decidida a utilização de uma estrutura de dados diferente da utilizada por eles, pois tal proposta não se adapta às características nem às finalidades previstas no projeto de pesquisa “*Estabelecimento de sistema automático e informatizado para análises de diversidade genética e genotipagem de acessos de cacauero *Theobroma cacao* L. com base em marcadores moleculares microssatélites*” (CORRÊA *et al.*, 2002), no qual este trabalho se insere, visando ter como diferença essencial otimização dos dados.

Visando avaliar a eficiência da estrutura de dados proposta foi considerado como objetivo específico construir um banco de dados relacional baseado nesta estrutura, e o povoamento parcial com dados provenientes do projeto “Diversidade genética de cacauero (*Theobroma cacao* L.) resistente à vassoura-de-

bruxa com base em marcadores RAPD e microssatélites” (LEAL, 2004) para teste de integridade e de coerência das consultas.

Desta forma, evita-se o armazenamento redundante e torna disponível um formato padrão que facilita a recuperação, re-análise e transferência de dados entre o banco de dados e os programas aplicativos.

A implantação de um sistema de informação centrado num banco de dados relacional, especificamente projetado para armazenar e disponibilizar, via WEB, informações fenotípicas e genotípicas de plantas, a fim de subsidiar pesquisadores e produtores envolvidos em programas de melhoramento genético assistidos por marcadores moleculares, constitui um passo importante no avanço destes programas.

2. REVISÃO BIBLIOGRÁFICA

Levando em conta o caráter interdisciplinar do trabalho, que vincula as Ciências Biológicas, em particular a Genética e a Biologia Molecular, e a Ciência da Computação, mais especificamente a Teoria dos Sistemas de Informação, a revisão bibliográfica foi dividida em duas fases: na primeira, focalizou-se o estabelecimento de conceitos e definições básicas da área de informática e tópicos referentes à arquitetura e desenvolvimento de software (itens 2.1 e 2.2); na segunda, foi estudado o problema do melhoramento genético baseado em marcadores moleculares (item 2.3).

2.1 Conceitos Básicos de Informática

2.1.1 Informática

A informática se ocupa do processamento e transmissão de informações em forma de dados. Significa “informação automática” sendo, portanto, uma forma de substituir o tratamento manual da informação pelo seu processamento por uma máquina (o computador).

A informática baseia-se em três tipos de recursos principais ou pilares: os equipamentos (hardware), os programas (software) e os recursos humanos capacitados para sua utilização (humanware).

2.1.1.1 Hardware

O hardware é um termo relacionado com os sistemas de processamento de dados, referindo-se ao equipamento físico, ou seja, aos componentes mecânicos, magnéticos ou eletrônicos. Inclui as unidades de processamento (computadores) e os equipamentos e dispositivos de transmissão de dados (distribuidores - concentradores, roteadores, cabos, conectores, etc.), isto é, tudo o que é tangível pelo operador do sistema. O hardware executa os programas (ou softwares), os quais traduzem as instruções que são enviadas para o computador em uma linguagem compreensível por este.

2.1.1.2 Software

Software de um sistema é o conjunto de elementos lógicos necessários à realização das tarefas requeridas ou a parte lógica que dota o equipamento físico (hardware) de capacidade para realizar uma tarefa específica de um determinado trabalho. Software, portanto, é uma sentença escrita em uma linguagem computável, para a qual existe uma máquina capaz de interpretá-la. A sentença (software ou programa) é composta por uma seqüência de instruções (comandos), declarações de dados e procedimentos, armazenáveis em meio digital. Ao interpretar o software, a máquina computável é direcionada à realização de tarefas especificamente planejadas, para as quais o software foi projetado. O software pode ser dividido em dois grandes grupos de programas: os básicos e os aplicativos.

2.1.1.2.1 Softwares Básicos

Os softwares básicos dividem-se em três sub-grupos: sistema operacional, ambiente operacional e tradutores de linguagem.

A. Sistema Operacional: Sem um Sistema Operacional (SO) nenhum hardware torna-se utilizável pelo usuário, porque ele é um conjunto de programas responsável pela interface primária entre o microcomputador e o usuário, atuando diretamente sobre o hardware do computador. O SO executa as mais diversas tarefas tais como o controle de entrada e saída (input/output), leitura/escrita em disco ou HD, escrita

em monitores e controle de todo o uso da memória dividindo-a em várias partes para cada tipo de aplicação.

Os sistemas operacionais mais utilizados são o Windows da Microsoft Corporation, o MacOS da Machintosh e os da família do Unix que foi desenvolvido pela BELLABS. A preferência por sistemas baseados em Unix deve-se ao fato de que tais sistemas são normalmente mais confiáveis, gerenciam melhor o trabalho, mesmo com grandes quantidades de dados, quando comparados a outros sistemas, e algumas de suas variantes, como o Linux, possuem código aberto (ou seja, pode ser visto e livremente modificado de acordo com as necessidades de quem o utiliza) e são distribuídas gratuitamente (PROSDOCIMI *et al.*, 2002; GLEIDSON, 2002).

B. Ambiente Operacional: Adiciona recursos ao sistema operacional como por exemplo uma interface gráfica com o usuário e com os dispositivos de rede, assim como com os múltiplos programas aplicativos para gerenciamento do computador e dos dados nele armazenados. Pode-se dizer que interfaces são os softwares que dão forma à interação entre o usuário e o computador. Atuam como uma espécie de tradutor, mediando as duas partes, tornando uma sensível à outra.

C. Tradutores de linguagem: São programas que permitem que os computadores executem programas não escritos em linguagem de máquina.

O processo de tradução pode ser realizado por dois tipos de tradutores: os compiladores* e os interpretadores*. Os dois tradutores aceitam como entrada um programa em linguagem de alto nível (fonte, compreendida pelo programador) e produzem como saída um programa em linguagem de máquina (objeto, compreendido pelo computador). A compilação é ideal para quando o programa estiver na sua forma definitiva ou for reutilizado com frequência. Um mesmo programa fonte compilado é mais rápido na execução que o modo interpretado. A grande desvantagem do programa objeto é não ser possível modificá-lo. Para tal é necessário retomar o código fonte, fazer as alterações e compilar tudo de novo. Ideal para comercialização de programas.

2.1.1.2.2 Softwares Aplicativos

Os programas aplicativos são construídos para a solução de um determinado problema, ou para a execução de uma determinada função. Por exemplo, o MS Word ® é um processador de textos e, como tal, tem ferramentas próprias para

edição e formatação de textos, desde uma simples carta até um elaborado jornal (Tipos de Softwares de Computadores, 2004). Os aplicativos estabelecem uma comunicação direta com o usuário (homem ou outro aplicativo em outro computador) via dispositivos de interface (mouse, teclado, monitor, placa de rede, etc.) e se comunicam com o hardware através do sistema operacional.

2.1.2 Linguagens

Os softwares, independente de sua classificação, são escritos com linguagens de programação específicas para determinadas finalidades.

Toda linguagem é formada por um conjunto de símbolos (palavras) organizados de acordo com um conjunto finito de regras de sintaxe que devem ser obedecidas para poder gerar sentenças com sentido não ambíguo nesse contexto.

Podemos classificar três tipos de linguagens segundo os participantes na troca de informações:

- Idiomas, entre seres humanos;
- Protocolos* entre computadores;
- Linguagem de programação, entre humanos e computadores.

As linguagens de programação permitem a formulação de instruções a um computador, ou seja, são utilizadas para escrever os códigos fonte dos programas de computador.

2.1.3 Sistema de Informação

Sistema é um conjunto de elementos interdependentes que interagem formando um todo unitário de tal forma que sejam atingidos os objetivos para os quais foi desenhado (YOURDON, 1990; REZENDE e ABREU, 2000).

O Sistema de Informação (SI) é definido por LAUDON e LAUDON (1999) como um conjunto de software, hardware e recursos humanos (componentes da tecnologia da informação e seus recursos integrados) que coleta, transforma e dissemina informações em uma organização (REZENDE e ABREU, 2000). A informação é o conhecimento explícito sobre fatos, regras, acontecimentos, etc. Os dados são representações codificadas da informação, utilizando um sistema de símbolos. Dados formatados e textos fazem parte integrante dos sistemas de informação.

2.1.3.1 Características e Finalidades dos Sistemas de Informação

A finalidade principal dos Sistemas de Informação é a de apoiar o processo de tomada de decisões em situações de decisão semi-estruturada, ou seja, onde existem diversas opções para a resolução de um problema e se deseja escolher a mais adequada (MARQUES e MARIN, 2002). Yourdon (1990) observa que os Sistemas de Informação não tomam decisões por eles mesmos, mas auxiliam na tomada de decisões humanas inteligentes e bem informadas, sobre vários aspectos de uma dada operação.

Para Schreiner Jr. (2001) os sistemas de informação apresentam, dentre outras, as seguintes características:

- Possibilitam o gerenciamento de dados e processos;
- Facilitam o cruzamento de informações seguindo critérios de seleção pré-definidos;
- Combinam modelos ou técnicas analíticas com as funções tradicionais do processamento de dados;

e devem:

- ser interativos e fáceis de usar;
- possuir interfaces amigáveis; e
- incorporar outras atividades, além do armazenamento de dados, como meios mais eficientes de distribuição dos resultados.

Os SI podem também:

- executar diversas análises matemáticas e estatísticas sobre os dados;
- apresentar informações sob várias formas gráficas (tabelas e diagramas); e
- fornecer informações relevantes em formato adequado, como relatórios convencionais, para que o gerente da instituição ou o coordenador do projeto, possa entender e acompanhar o processo gerador das informações.

A literatura especializada consultada aborda amplamente os SI corporativos, mas não dedica a devida atenção aos sistemas para apoio a atividades de pesquisa e desenvolvimento como é o caso considerado neste trabalho. Por isso considera-se conveniente abordar a classificação dos SI empresariais para traçarmos um paralelo entre eles.

Segundo O'Brien (2001), de acordo com sua finalidade, os Sistemas de

Informação Empresariais podem-se classificar em dois grupos distintos: os sistemas de apoio ao trabalho interno e os sistemas de apoio à gerência, que se sub-classificam cada um deles em outros três tipos de sistemas, como mostra a Figura 1.

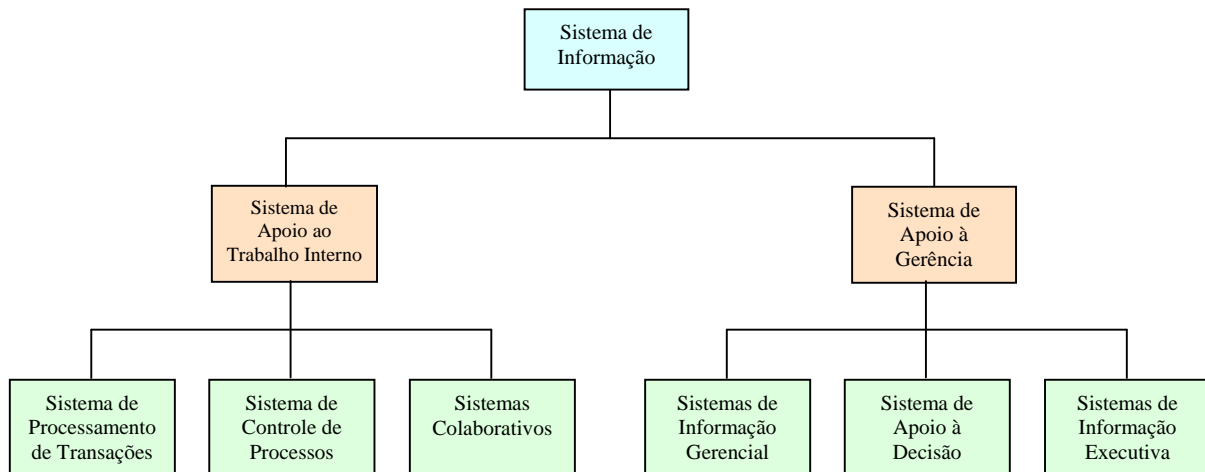


Figura 1. Classificação de SI conforme o propósito (adaptado de O´BRIEN, 2001).

Essa classificação se adequa aos processos de negócios das empresas do setor produtivo e focaliza as atividades produtiva e comercial. Os Sistemas de Apoio ao Trabalho Interno, se dividem em:

- Sistema de Processamento de Transações: registra e controla transações de inserção, manipulação e exclusão;
- Sistema de Controle de Processos: controla o setor de produção e a cadeia de suprimentos;
- Sistema Colaborativo: possibilita a colaboração entre os sistemas de processamento de transações e controle de processos e controla outros setores, como o de recursos humanos e o financeiro.

Os Sistemas de Apoio à Gerência por sua vez se subdividem segundo sua finalidade em:

- Sistema de Informação Gerencial: monitora os setores de produção e comercialização;
- Sistema de Apoio a Tomada de Decisão: manipula e relaciona índices de forte impacto no desempenho da empresa;
- Sistema de Informação Executiva: monitora balanços e a saúde financeira da empresa.

Um Sistema de Informação para apoiar Projetos de Pesquisa deve possuir funcionalidades tanto para facilitar o trabalho interno do grupo de pesquisa, no sentido de disponibilizar uma ferramenta de uso comum para armazenamento e consulta dos dados gerados na pesquisa, quanto para permitir o acompanhamento e controle do avanço das pesquisas por parte da coordenação do projeto ou programa, subsidiando, caso necessário, o processo de tomada de decisão.

2.1.3.2 Sistema de Informação baseado na WEB

Sistema de Informação baseado na WEB é aquele que usa as funcionalidades oferecidas pela tecnologia WEB como, por exemplo, arquitetura de comunicação capaz de suportar um grande número de acessos, questões de segurança (quando

se tratar de uma infra-estrutura de comunicação pública) e interligação com outros sistemas existentes (ZANETI, 2003).

Como mostra a Figura 2, a troca de informações entre usuários e um sistema WEB é bidirecional, similar ao que ocorre em outros Sistemas de Informação.



Figura 2. Funcionamento de Sistemas Baseados na WEB.

2.2 Arquitetura e Desenvolvimento de Software

Por arquitetura de software entende-se o estudo da organização global dos sistemas de software bem como o relacionamento entre seus componentes, permitindo o entendimento dos mesmos, especialmente dos componentes que são consistentes ao longo do tempo e de implementações (SILVA FILHO, 2002).

2.2.1 Arquitetura dos Sistemas de Informação

A Arquitetura dos Sistemas de Informação é vista como o estabelecimento de um conjunto de elementos cuja

finalidade é proporcionar um mapeamento da organização para a qual o sistema é requerido, sobretudo no tocante aos elementos envolvidos com o processo de desenvolvimento e implantação do SI.

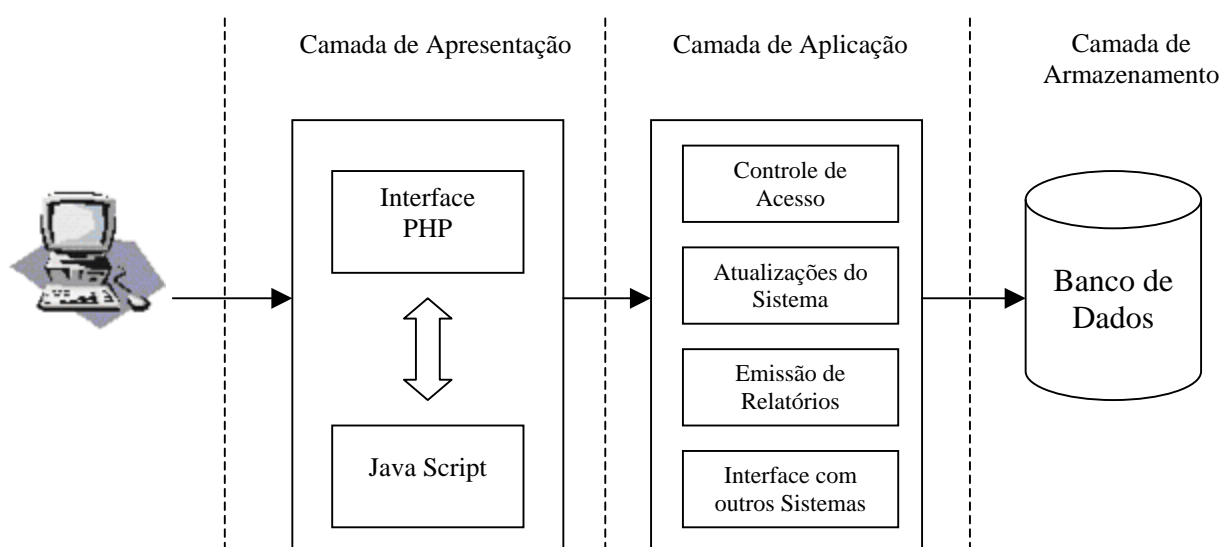


Figura 3. Componentes que formam a estrutura da arquitetura de um sistema de informação baseado na WEB, evidenciando as camadas do sistema.

Na Figura 3, temos: (i) a Camada de Apresentação composta pela interface gráfica com o usuário baseada em janelas, (ii) a Camada de Lógica de Aplicação que implementa as funcionalidades da aplicação e fornece serviço de apoio, tal como a interface com um BD e (iii) a Camada de Armazenamento (ou de Persistência), contendo um BD relacional. Estas camadas compõem a arquitetura geral de um SI, formada por cinco grupos de componentes como os descritos a seguir:

1. Banco de Dados (BD): São coleções de dados relacionados a um propósito específico, um conjunto de informações relacionadas entre si, referentes a um mesmo assunto e organizadas de maneira útil, com o

propósito de servir de base para que o usuário recupere informações, tire conclusões e tome decisões. Existem quatro tipos básicos de bancos de dados e a escolha por um determinado tipo interfere no resultado desejado para o armazenamento de dados (DATE, 2001).

a. Banco de dados de indexação de arquivos, o *flat file* (ou de arquivo simples), o qual não deve ser considerado um típico Sistema Gerenciador de Banco de Dados (SGBD), por ser nada mais que uma coleção de arquivos semelhantes ordenados, como se fosse o conhecido arquivo de metal, possuindo pastas etiquetadas e organizadas como é o arquivo físico (GIBAS e JAMBECK, 2002).

b. Bancos de dados relacionais (BDr). Nos bancos de dados relacionais as informações ficam distribuídas e organizadas em tabelas (linhas x colunas) inter-relacionadas. É o tipo mais utilizado de banco de dados (GIBAS e JAMBECK, 2002). A vantagem desta estrutura se dá porque estes bancos fornecem uma ótima forma de armazenar e acessar informações complexas e facilitam a consulta de dados, proporcionando diferentes visões dos mesmos.

c. Bancos de dados orientados a objetos. São bancos que agrupam as informações em objetos persistentes e são projetados para administrar interações simultâneas por vários clientes. A unidade de armazenamento, o objeto, é concebido como uma entidade abstrata que representa um objeto do mundo real, mais especificamente do domínio de aplicação do SI. Por exemplo nos programas de melhoramento genético de plantas, um objeto pode ser uma variedade da mesma. A esse objeto se atribuem as características e propriedades genotípicas, fenotípicas e agrônômicas, etc. O banco de dados orientado a objeto combina os benefícios e conceitos da orientação a objetos com a funcionalidade dos bancos de dados (KHOSHAFIAN, 1994).

d. Bancos híbridos (objeto-relacionais): Tipos de bancos de dados nos quais as informações estão organizadas em forma de tabelas (linhas x colunas) inter-relacionadas, porém, pode-se definir uma sub-estrutura de campos como se fossem objetos. Vejamos como exemplo duas tabelas de cadastros de *marcadores moleculares*: uma de *acesso* e uma de *loco do*

marcador. Os dados comuns destas tabelas (nome do acesso, data de inserção, data de atualização, etc.) podem ser definidos em forma de objeto, ou seja, define-se um objeto contendo os campos de nome do acesso, data de inserção e data de atualização e atribui-se este objeto às tabelas, definindo normalmente os demais campos ou campos específicos. Quando alguma alteração de estrutura for realizada nos campos-objeto, ela refletirá automaticamente nas tabelas em que foram utilizados (Documentação do PostgreSQL 7.4.1, 2005).

2. Sistema Gerenciador de Banco de Dados (SGBD): Este sistema é responsável por intermediar os processos de construção, manipulação e administração do banco de dados solicitados pelo Ambiente de Interação ou por outras aplicações. Após os dados estarem instalados no BD, o SGBD deve possibilitar o acesso às informações e a sua atualização, garantindo a segurança e a integridade do BD. É o programa que gerencia e disponibiliza os dados para consultas. Todos os grandes bancos de dados fazem uso da Linguagem de Consulta Estruturada – Structured Query Language (SQL). Alguns dos mais populares SGBD relacionais comerciais são o Oracle, Sybase, Informix, SQL Server da Microsoft e Db2 da IBM. Os dois principais SGBD relacionais de código aberto e gratuito (para S.O. Linux) são o MySQL e o PostgreSQL (GREENSPAN e BULGER, 2001). O MySQL é bastante utilizado em bancos de dados de tamanho pequeno a médio, por ser fácil de configurar e usar e permitir acesso veloz aos dados do banco. Para servir um conteúdo WEB ou criar um site institucional de tamanho moderado, MySQL tem o potencial necessário. O PostgreSQL também é um SGBD gratuito, com ferramentas muito poderosas, entretanto não é muito utilizado pela dificuldade no seu gerenciamento (GIBAS e JAMBECK, 2002; PRODOCIMI *et al.*, 2000).
3. Ambiente de Interação (AI): Com o desenvolvimento dos computadores, atualmente até os não especialistas podem utilizá-los com facilidade, exigindo cada vez menos conhecimentos técnicos em informática. O que possibilita uma relação sem grandes dificuldades entre homem e máquina chama-se de interface amigável. É este ambiente de interação que

propicia a interface homem-máquina, a comunicação entre ambos de forma a existir compreensão das necessidades mútuas. Um exemplo de ambiente de interação é uma página HTML, num portal WEB. Esta página é visualizada pelo usuário, que pode interagir com ela de forma a querer efetuar uma consulta de dados de um banco de dados. O usuário não visualiza o banco, apenas a página em HTML, na qual faz o requerimento desejado. As informações são capturadas e analisadas por um programa (ex. escrito em PHP) que organiza a consulta para que o SGBD entenda, recupere os dados desejados, que serão disponibilizadas em formato HTML, visualizado por um navegador de internet, de forma que o usuário compreenda.

4. Ambiente Aplicativo (AA): Formado pela lógica de aplicação, a qual é responsável por atender os requisitos funcionais do sistema e prover uma intermediação com um banco de dados através do SGBD e por processos de manipulação de dados acrescentados ao sistema de informação que fazem análise de dados primários e fornecem dados secundários relevantes para o domínio de aplicação do sistema.

5. Ambiente Operacional (AO): É composto por hardwares, softwares e pessoas que permitem que todos os componentes do ambiente de aplicação sejam integrados. Aí podem estar incluídos microcomputadores (em rede ou não), scanners de código de barra, estações de trabalho, outros softwares que colaboram, outros bancos de dados associados, etc.

O relacionamento entre os componentes e as camadas é mostrado no Quadro 1:

	Camada de Armazenamento	Camada de Aplicação	Camada de Interface
BD			
SGBD			
AI			
AA			
AO			

Quadro 1 – Relacionamento entre os componentes e as camadas que formam a arquitetura de sistemas de informação. A área hachurada indica a presença dos componentes nas camadas da arquitetura do Sistema.

2.2.2 Processo de Desenvolvimento de Software Orientado a Objeto

O desenvolvimento de software de média e alta complexidade não é possível sem a utilização de um enfoque metodológico apropriado que esteja de acordo com as diretrizes básicas da Engenharia de Software.

O ciclo de vida de um software é definido a partir dos diferentes estágios pelos que transita qualquer programa de computador, começando pela sua concepção ou definição como “solução” a um “problema” existente num domínio de aplicação, passando pela fase de construção até sua transição do local onde foi produzido para o local onde será utilizado. Após a implantação do software como parte integrante, seja ativa ou passiva, do processo que tinha problemas, começa o estágio de suporte pós-implantação (ou suporte ao cliente), na qual o software pode vir a sofrer modificações, corretivas visando a eliminação de erros não detectados durante sua construção ou adaptativas para se atender a novas exigências de uso e ou a novas tecnologias de hardware ou software básico. Contudo, pode chegar um ponto em que o software não assimile mais adaptações sem comprometer suas funcionalidades básicas e tenha que ser substituído por outro mais novo. Esta visão de desenvolvimento de software como um ciclo de vida é uma visão gerencial do processo (VASCONCELOS, A. M. L., 2004; PAULA Filho, W. P., 2003)

A metodologia mais aceita atualmente para um projeto de desenvolvimento de software considera uma seqüência de sete atividades principais:

1. Levantamento de Requisitos;
2. Análise;
3. Projeto;
4. Implementação ou codificação;
5. Testes Alfa;
6. Testes Beta;
7. Manutenção.

Estas atividades contemplam uma visão técnica do desenvolvimento de software, em que não se considera a fase de retirada ou “morte” do software.

O Quadro 2 proporciona a visão que cada grupo (o de gerência e o técnico) tem a respeito do desenvolvimento de software.

VISÃO GERENCIAL	VISÃO TÉCNICA
Ciclo de Vida	Ciclo de Desenvolvimento de Software
Concepção / Definição	Levantamento de Requisitos e Análise
Construção	Projeto Implementação Teste α
Implantação	Teste β
Suporte ao cliente	Manutenção } Corretiva Evolutiva
Retirada (“morte” do software)	—

Quadro 2. Visão gerencial e visão técnica de desenvolvimento de software. Concepções diferentes para um mesmo processo.

Para o desenvolvimento respaldado na metodologia de visão técnica, as duas primeiras atividades são realizadas no estágio de concepção, as próximas três atividades no estágio de construção e as duas últimas conformam o fluxo de trabalho característico das fases de transição e suporte pós-implantação.

O Levantamento de Requisitos permite estabelecer um canal de comunicação entre o desenvolvedor e o cliente/usuário visando a definição das funções a serem executadas pelo software dentro do processo de negócio assim como outras características desejadas, de desempenho, estéticas, de segurança, etc. Os principais resultados (chamados de artefatos na Engenharia de Software) desta atividade são:

1. Diagrama do Contexto, onde se modela o processo de negócio e se identificam os pontos onde o software participa;
2. Especificação Formal de Requisitos, onde se especificam as funções do software e outras informações suplementares;
3. Modelo de Casos de Uso, onde se identificam os usuários do sistema, chamados comumente de atores, e se associam com as principais funções do software (chamados de casos de uso). Um caso de uso representa uma funcionalidade completa do software contando de um fluxo principal e sub-fluxos alternativos de acordo com a lógica do programa;
4. Protótipo da Interface, onde se estabelecem aspectos relacionados com o layout, apresentação, interatividade, etc.;
5. Glossário, onde se estabelecem as definições de termos próprios do domínio de aplicação, para consulta pelos desenvolvedores.

A Análise é a atividade que objetiva modelar “o que” e “como” o software irá fazer, independente da tecnologia que será utilizada para sua construção posterior. No paradigma Orientado a Objeto (LARMAN, 1999) a tarefa de modelagem consiste na identificação de estruturas conceituais chamadas de objetos que descrevem no seu conjunto o domínio de aplicação do software. Cada objeto representa uma parte ativa do processo, isto é, possui propriedades ou atributos (considerados dados relevantes para o processo) e as funções necessárias para adquiri-los, manipulá-los, transformá-los, exportá-los, etc. Os objetos podem aparecer em determinado estágio da execução do software e desaparecer em outro, quando já sua função no processo não for mais necessária. Neste enfoque os casos de uso que representam as funcionalidades do software se implementam através da colaboração dinâmica de diversos objetos que se comunicam entre si. A interação entre dois objetos é considerada uma troca de mensagens. Os principais resultados desta fase são:

1. Diagrama de Classes, onde se representam as diversas classes de objetos identificados e a forma como se comunicam entre si;
2. Diagrama de Entidade-Relacionamento (DER), onde se representam as diferentes entidades conceituais identificadas no domínio e se especificam os

*relacionamentos entre elas. Uma entidade pode ser caracterizada como um objeto de dados persistentes, i.e, uma entidade do domínio de aplicação do software, cujos dados devem ser armazenados de forma permanente.*²

A atividade de Projeto permite a construção de um modelo físico implementável do software. Nesta fase se define a arquitetura do software e a tecnologia de construção, i.e, a plataforma operacional e as linguagens de programação a serem utilizadas levando em conta as características estabelecidas nas fases anteriores. Os principais resultados são:

- 1. Diagrama de Componentes, onde se representa a arquitetura do software, identificando-se seus módulos e as camadas de interface entre os mesmos;*
- 2. Diagrama de Colaboração de Classes, onde se representam as diversas classes que participam na implementação de cada caso de uso e sua interação dinâmica;*
- 3. Diagrama de Instalação/Distribuição onde se especifica a forma de instalação do software, indicando em que máquinas e como devem ser instalados os diversos módulos e pacotes da distribuição;*
- 4. Projeto Físico do Banco de Dados onde se definem estruturas de dados dos repositórios permanentes;*

A atividade de Implementação, também conhecida como codificação, visa a transformação do modelo físico do software em código de computador, utilizando linguagens de programação. Nesta fase se devem construir também as interfaces entre os diversos módulos ou componentes, como por exemplo entre o banco de dados e a interface com o usuário no caso de sistemas de informação. Os principais resultados são:

- 1. Códigos fontes dos componentes do software;*
- 2. Códigos executáveis e scripts do software.*

A atividade de Testes Alfa se divide em testes de componentes e testes de integração. Os testes de componentes se realizam geralmente durante o próprio processo de codificação do software servindo de mecanismo de realimentação ao

² O DER é um diagrama de Análise Estruturada e, embora não se prevê DER para UML e o paradigma escolhido para a estruturação do sistema seja do tipo Orientado a Objeto, este diagrama foi utilizado para melhor compreensão inicial da Análise Orientada a Objeto, por parte da autora.

programador. Os testes visam validar o cumprimento dos requisitos previamente definidos e o fluxo correto dos casos de uso implementados. Os testes de integração, como seu nome indica, visam avaliar o funcionamento harmônico dos diversos componentes do software uma vez integrados. Em caso de detecção de erro o processo é detido e se determina a causa e a forma mais apropriada para corrigi-lo. Muitas vezes nesta fase se identificam erros não apenas de codificação, mas de projeto ou análise, forçando a modificação dos artefatos correspondentes e o acompanhamento das modificações em cascata. Os principais resultados são:

1. *Relatório de testes de componentes;*
2. *Relatório de testes de integração*

Os Testes Beta são as atividades associadas à avaliação preliminar feita pelo usuário no ambiente real ou simulado de aplicação do software. De acordo com seus resultados o software será implantado ou revisto para satisfazer as expectativas do cliente. O tempo necessário para esta atividade não pode ser facilmente estimado, e depende da complexidade do software, da natureza do processo que está sendo informatizado, da experiência dos usuários-chave encarregados dos testes, entre outros. Os principais resultados são:

1. *Formulário de Relatório de Testes;*
2. *Termo de Aceitação.*

A atividade de Manutenção acontece uma vez que o software foi implantado nas instalações do usuário e os usuários foram devidamente treinados. A atividade de manutenção pode ser corretiva, solucionando problemas detectados após a instalação ou evolutiva, conduzindo a novas versões do mesmo software. O tempo necessário para esta atividade também não pode ser estimado facilmente. Os principais resultados são:

1. *Versões curadas do software;*
2. *Novas versões Beta.*

2.2.2.1 Modelos de Ciclos de Desenvolvimento

Existem vários modelos de ciclo de vida, a depender da metodologia que se pretende seguir, i.e, da forma como as fases principais (levantamento de requisitos, análise, projeto, implementação e testes) são executadas.

Os principais modelos de ciclo de vida existentes são: modelo cascata, modelo de desenvolvimento evolucionário e modelos iterativos* (espiral e incremental) (VASCONCELOS, 2004).

2.2.2.1.1 Modelo Cascata

Neste modelo o desenvolvimento do software se faz com execução sistemática e serial de suas fases, uma após a outra. Porém, com os sofisticados sistemas de software da atualidade, não é mais possível definir primeiramente todo o problema, desenhar a solução inteira, desenvolver o software e depois testar o produto no final do projeto. Por isso que na prática o que se observa é o desenvolvimento do software de forma dinâmica, com repetições cíclicas entre elas, visando detectar a tempo erros na execução de cada fase e corrigi-los.

2.2.2.1.2 Modelo de Desenvolvimento

Evolucionário

Este modelo subdivide-se em outros dois modelos: o de Programação Exploratória e o de Prototipação Descartável.

O Modelo de Programação Exploratória tem por objetivo agilizar o desenvolvimento da primeira versão do sistema, de maneira que o escopo deste não tem definição clara, pois é especificado paralelamente ao desenvolvimento.

Já o Modelo de Prototipação Descartável objetiva a compreensão dos requisitos do sistema. Assim como no Modelo de Programação Exploratória desenvolve-se um protótipo para experimentação/teste porém, neste modelo, o protótipo é descartado depois e supõe-se a utilização de um outro modelo de ciclo de vida para de fato implementar o software desejado.

2.2.2.1.3 Modelos Iterativos

Seguir um planejamento iterativo é interessante, pois nos permite um aumento significativo no entendimento dos problemas através de refinamentos sucessivos. A divisão do projeto em múltiplas iterações torna mais produtivo todo o ciclo de desenvolvimento reduzindo significativamente os riscos do projeto através da demonstração freqüente do progresso e releases (partes liberadas do software) executáveis que possibilitam o envolvimento e suporte do usuário final. A proposta deste modelo é tornar o desenvolvimento de software dinâmico, sendo contrário à seqüência de atividades estáticas seguida por um modelo serial.

Dentro dos modelos iterativos há dois tipos de enfoque: Desenvolvimento Espiral e Desenvolvimento Incremental.

No Modelo de Desenvolvimento Espiral existem quatro fases de atividades, como pode ser visto na Figura 4 sendo cada fase dinâmica e interdependente (representada pelos algarismos), como se simulassem quadrantes de um espiral: a) determinação dos objetivos, alternativas e restrições; b) análise das alternativas e identificação e/ou resolução de riscos; c) desenvolvimento e validação da versão atual do produto e; d) planejamento (fase em que uma nova espiral é planejada em vista do resultado obtido nas outras três fases). A, B e C marcam o início de cada ciclo e a cada etapa pode culminar na concepção de novas características, dependendo da necessidade de um novo ciclo.

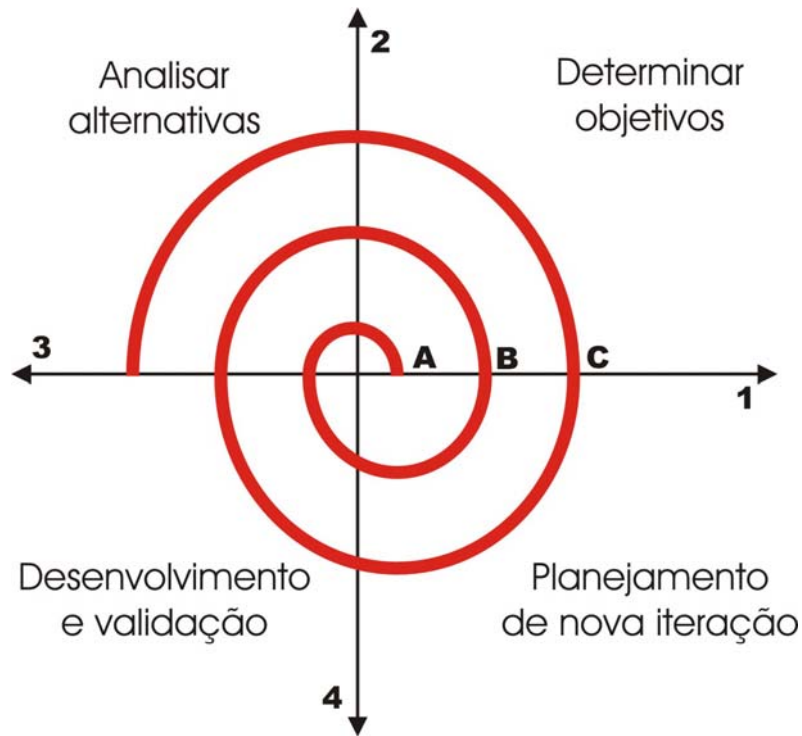


Figura 4. Etapas de um ciclo do modelo espiral.

O Modelo de Desenvolvimento Incremental, embora tenha um enfoque iterativo, difere do espiral na entrega do sistema: enquanto que no modelo espiral, após todas as iterações necessárias, o sistema é entregue completo, num Modelo incremental o sistema é entregue por partes, e cada parte representa um incremento para a parte anterior, seguindo a idéia que requisitos de mais prioridade são vistos nos ciclos iniciais. Assim, pode-se ter o retorno do usuário, como se os primeiros incrementos fossem protótipos do sistema, para novos requisitos dos incrementos finais deste sistema (Figura 5).

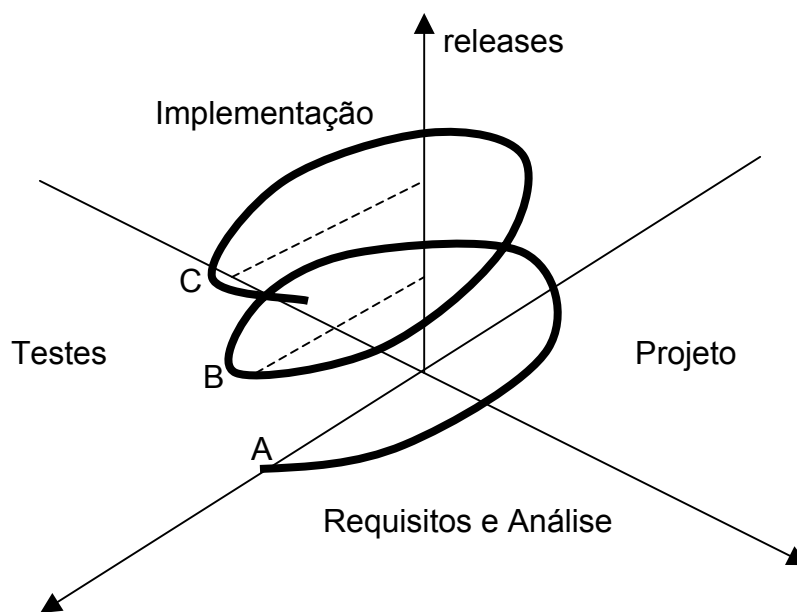


Figura 5. Modelo Iterativo Incremental. A, B e C representam o início da 1ª, 2ª e 3ª iterações, respectivamente.

Há vários modelos de desenvolvimento sendo atribuição do desenvolvedor a escolha do modelo que melhor se adequa às características do sistema a ser desenvolvido.

2.3 Caracterização dos Programas de Melhoramento Genético Baseados em Marcadores Moleculares

Os principais motivos para desenvolver programas de melhoramento genético, animal ou vegetal, são: aumentar a resistência a doenças, aumentar a produtividade e melhorar a qualidade do produto.

Os programas de melhoramento consistem na produção de variabilidade genética na população seguida pela seleção dos genótipos desejáveis.

O primeiro e mais importante passo no desenvolvimento de um programa de melhoramento genético consiste em definir seus objetivos, os quais devem considerar a possibilidade de se obter resposta à seleção, dada pelos valores de herdabilidade e variabilidade das características, de forma a possibilitar lucro financeiro ao produtor, bem como retorno ao investimento feito em melhoramento da espécie. Definidos os objetivos do programa, é necessária a formação de uma coleção de genótipos (SONG, 1980; ALCAZAR, 1983 e CAMARGO *et al.*, 1974),

onde as espécies estudadas sejam representadas por vários acessos, os quais precisam ser catalogados e avaliados, visando obter conhecimento sobre a diversidade genética existente entre e dentro das populações (STEFENON, 2003). Este conhecimento adquirido também acaba permitindo o intercâmbio entre instituições de pesquisa.

Em plantas existem duas maneiras básicas de se constituir um programa de melhoramento genético: a forma convencional e a não-convencional.

O tipo de melhoramento convencional é constituído de três estágios: a criação da variação, a seleção de variantes com características desejáveis e testes de campo para confirmar se as variedades selecionadas têm as características requeridas. Ele pode basear-se em hibridação intra ou inter-específica, ou seja, os indivíduos em estudo (plantas ou animais) são cruzados entre si, independentemente de serem de mesma espécie (cruzamento intra-específico) ou de espécies diferentes (cruzamento inter-específico), seus descendentes são avaliados e alguns desses descendentes são selecionados pelos melhoristas.

O melhoramento não-convencional pode ser feito com auxílio de técnicas como a micropropagação, o resgate de embriões, a transformação genética (através da obtenção de plantas transgênicas), a produção de mutantes por bombardeamento radioativo, indução de mutação associada às técnicas *in vitro* e a fusão de protoplastos (BORÉM, 1999).

Mais recentemente, o estudo da diversidade genética por marcadores moleculares, o seqüenciamento dos genomas com uso de máquinas de seqüenciamento automático de DNA e o mapeamento dos genes de interesse passaram a ser amplamente utilizados em estratégias de melhoramento (LEAL *et al.*, 2004). Todas essas técnicas necessitam de sistemas de informação e outros recursos de informática, tendo em vista o grande volume de dados produzidos e que precisam ser analisados de maneira mais rápida, precisa e eficiente.

A genética Mendeliana associada à genética quantitativa e aos métodos estatísticos constituem as bases do melhoramento genético convencional. A importância e contribuição dela são incontestáveis no melhoramento genético de plantas. Porém, sabe-se que o melhoramento genético convencional é um processo demorado³, principalmente se considerarmos plantas perenes, e é cada vez mais trabalhoso e mais dispendioso obter ganhos. Contudo, constata-se um grande avanço no processamento de dados (novos modelos estatísticos e computacionais) e facilidade de intercâmbio de informações, germoplasma e uma maior interdisciplinaridade com outras ciências (Agronomia,

³ Em geral de 5 a 10 anos, sendo que, em alguns casos, pode chegar a mais de 30 anos para se desenvolver uma variedade, como é o caso de espécies florestais de clima temperado.

Fitopatologia, Entomologia, etc.), proporcionando maior produção e ganho de conhecimentos.

Os marcadores moleculares aliados às técnicas de engenharia genética e de seqüenciamento de genomas têm potencial para revolucionar o melhoramento genético de organismos, permitindo aos melhoristas produzir genótipos com maior produtividade e tolerância a estresses bióticos e abióticos, mais eficientemente, mais rápido e mais econômico do que o melhoramento convencional.

Banco de Germoplasma é uma coleção de espécies da máxima importância onde estão armazenados os recursos genéticos de uma espécie. Este depósito de informações genéticas possibilita a abertura de novas oportunidades no crescente mundo da bioengenharia, aproveitando a enorme biodiversidade existente no território brasileiro. A variabilidade genética disponível utilizada tem ocorrido naturalmente e bancos de germoplasma existem para preservar esta variabilidade. Assim, estes bancos desempenham um papel extremamente importante nos programas de recursos genéticos e melhoramento; pois representam o repositório da variabilidade genética a disposição do melhorista (ALVES, 2002, TODA FRUTA, 2004)

Os cruzamentos das espécies estudadas permitem a produção de novas e desejáveis recombinações de genes, eventos propícios para a formação de variabilidade genética. Mas quando o banco de germoplasma não contém mais as combinações gênicas nem as mutações* espontâneas desejáveis, o uso das técnicas de indução de mutação é uma alternativa, pois permite a geração da variabilidade desejada e tem acelerado o processo no melhoramento de espécies cultivadas (MALUSZYNSKI *et al.*, 1995, DONINI e SONNINO, 1998).

Mas por que a variabilidade das espécies é um objetivo tão importante de ser alcançado, a ponto de justificar programas de melhoramento genético?

A diversidade genética das espécies é crucial para manter a capacidade natural de responder às mudanças climáticas e a todos os tipos de estresses bióticos e abióticos, sendo necessária para a obtenção de ganho genético. Em geral, quando outros fatores se mantêm constantes, o ganho genético (ex. aumento na produtividade) é diretamente proporcional à diversidade genética na população. Porém, o que se observa recentemente é uma perda acentuada desta diversidade. Esta perda de materiais genéticos, denominada erosão genética, significa a perda da variabilidade necessária para o melhoramento genético das plantas, tornando-as mais vulneráveis a fatores adversos. Quanto mais há uniformidade genética, ou seja, não há variação nos genótipos, mais há comprometimento de sua

resistência às mudanças climáticas e a resistência e/ou tolerância a diferentes tipos de estresses, devido a esta falta de variabilidade pois, se um indivíduo é atacado por uma determinada doença, por exemplo, possivelmente todos os exemplares da área serão atingidos.

Um dos fatores que tem contribuído para o agravamento do Quadro produtivo de espécies em plantações é a vulnerabilidade* genética dos cultivares ao ataque de patógenos e à rápida disseminação devido à homogeneidade das plantas. No caso do cacauero, isto se deve, principalmente, à forma de propagação vegetativa utilizada, que favorece a disseminação de doenças (material original de genótipos similares entre si e susceptíveis) e torna os cultivos muito uniformes geneticamente e, assim, mais susceptíveis às doenças.

A caracterização de cultivares com marcadores moleculares, quando feita sistematicamente e com técnicas padronizadas, pode levar a obtenção de um banco de dados valioso para um programa de melhoramento. Esses dados poderão ser utilizados para inúmeros objetivos, incluindo: a verificação de genealogias não confiáveis ou com falta de informação (LEE *et al.*, 1990); avaliação da variabilidade existente no programa e grau de diversidade genética (O'DONOUGHUE *et al.*, 1994; CAO e OARD, 1997); auxílio no planejamento de cruzamentos, especialmente no caso de genótipos elites e aparentados (LEE, 1995); definição de grupos heteróticos (HONGSTRAKUL *et al.*, 1997); entre outros. No caso de descritores morfológicos, isso não é possível de ser feito com a precisão requerida pelos melhoristas.

A biologia molecular, em particular com a aplicação dos marcadores moleculares, associados às informações de seqüenciamento e de mapeamento de genomas, começa a gerar informações importantes sobre os fatores relativos à biologia e à arquitetura genética (número de locos controlando uma característica, número de alelos segregando por loco, efeitos causados pela substituição de alelos, regulação gênica) das características quantitativas de importância

agronômica. O emprego de marcadores também poderá ser útil no estabelecimento da identidade dos acessos, permitindo detectar erros de identificação no campo, além de complementar a caracterização realizada com caracteres morfológicos e isoenzimáticos.

Marcadores moleculares, mais especificamente os de DNA, são passíveis de serem detectados e utilizados como procedimentos de melhoramento, podendo ser utilizados em praticamente todas as fases de um programa de melhoramento: na análise de diversidade, na escolha de progenitores, no mapeamento de regiões genômicas, no melhor entendimento da herança de caracteres, em *fingerprint*, dentre outras. Os diferentes tipos de marcadores moleculares hoje disponíveis, diferenciam-se pela tecnologia utilizada para revelar variabilidade em nível de DNA. As características morfológicas básicas são influenciadas pelo ambiente. Assim sendo, a partir do momento que marcadores moleculares permitem detectar a variabilidade em nível de DNA, ou seja, com mínima influência ambiental, eles podem ser considerados uma ótima ferramenta para este fim.

2.3.1 Tipos de Marcadores Moleculares

Vários tipos de marcadores moleculares de DNA estão hoje disponíveis diferenciando-se entre si quanto à sua habilidade em detectar polimorfismo* entre indivíduos geneticamente relacionados (que varia muito com o tipo de marcador utilizado para cada espécie), custo de aplicação, facilidade de uso e consistência de resultados.

Os principais tipos de marcadores moleculares podem ser classificados em dois grupos, conforme a metodologia utilizada para identificá-los: hibridização ou amplificação de DNA. Entre os identificados por hibridização estão os marcadores RFLP (“Restriction Fragment Length Polymorphism”) e minissatélites ou locos VNTR (“Variable Number of Tandem

Repeats”). Já aqueles revelados por amplificação incluem os marcadores do tipo: RAPD (“Random Amplified Polymorphic DNA”); SCAR (“Sequence Characterized Amplified Regions”) ou ASA (Amplified Specific Amplicon); AFLP (Amplified Fragment Length Polymorphism); e microssatélite (ou SSR - “Simple Sequence Repeats”). As tecnologias de marcadores moleculares estão evoluindo rapidamente e modificações já existem para algumas das técnicas acima mencionadas.

2.3.1.1 RFLP

A técnica de RFLP é complexa, mais demorada que as outras técnicas para obtenção de resultados; de custo relativamente alto; e tem revelado um grau de polimorfismo de intermediário a baixo, conforme a espécie. Mesmo assim, os RFLPs têm sido utilizados em um grande número de estudos de caracterização de cultivares (GEBHARDT *et al.*, 1989; O’DONOUGHUE *et al.*, 1994; AUTRIQUE *et al.*, 1996; N’GORAN *et al.*, 1994) . Isso tem sido devido principalmente à sua alta consistência e repetibilidade na obtenção dos resultados.

2.3.1.2 Minissatélites

Os minissatélites ou locos VTNR são seqüências repetitivas de DNA, adjacentes e em número variável (JEFFREYS *et al.*, 1985). Essa técnica é similar à de RFLP, variando

basicamente o tipo de sonda utilizado, e apresentado as vantagens e desvantagens já apresentadas para a técnica anterior. Uma vantagem adicional dos minissatélites é o alto grau de polimorfismo apresentado, decorrente da variação na distribuição dos sítios de restrição, das sondas utilizadas, e do número e tipos das seqüências repetitivas.

2.3.1.3 RAPD

A técnica de RAPD, dentre as apresentadas, é a de menor custo, número de etapas, e tempo para obter os resultados; e é fácil de implementar. Essa, contudo, tem a desvantagem de ser de repetibilidade baixa e pouco consistente de um laboratório para o outro, o que dificulta a comparação de dados obtidos em diferentes locais. Assim, cuidados devem ser tomados na padronização da técnica no laboratório para a caracterização de cultivares. O nível de polimorfismo obtido com RAPDs varia grandemente com a espécie em questão, e tem sido utilizada com sucesso na caracterização de variedades de cevada (TINKER *et al.*, 1993; PENNER *et al.*, 1998), e arroz (MACKILL, 1995), de cacau (FALEIRO *et al.*, 2001a; YAMADA *et al.*, 2001; N'GORAN *et al.*, 1994) entre outras.

2.3.1.4 SCAR

Marcadores SCAR são amplificados com *primers*^{*} específicos, desenvolvidos com base em seqüências já mapeadas ou caracterizadas (PARAN e MICHELMORE, 1993). Muitos desses *primers* são obtidos da conversão de marcadores RAPD em SCAR. Essa conversão em geral resulta na diminuição do nível de polimorfismo obtido por SCAR. Contudo, isso pode ser amenizado com a digestão com enzimas de restrição dos produtos amplificados; e com o seqüenciamento das bandas monomórficas e subsequente desenvolvimento de *primers* que amplificam seqüências mais variáveis entre os genótipos. A técnica de SCAR é muito semelhante a de RAPD, com a vantagem de ser mais consistente e desvantagem de envolver o desenvolvimento de *primers*, o que eleva o custo.

2.3.1.5 AFLP

A técnica de AFLP possui grande capacidade para detecção de variabilidade genética e uso em caracterização de cultivares. De fato, tem sido utilizado com sucesso em girassol (HONGTRAKUL *et al.*, 1997) e em cacau (FALEIRO *et al.*, 2001a) para essa finalidade. Entre as vantagens do uso desta estão o alto grau de polimorfismo e o mais alto número de marcadores obtidos por gel analisado. AFLP é a mais elaborada das técnicas de PCR, necessita gel de poliacrilamida para resolução dos fragmentos e é protegida por patente.

2.3.1.6 Microssatélites

Os microssatélites (ou SSR – Simple Sequence Repeats), que envolvem o desenvolvimento de *primers* específicos, é um processo elaborado e caro. Por isso, a aplicação rotineira de marcadores microssatélites requer que a relação custo/benefício seja considerada pois, as limitações na utilização de SSR são, geralmente, atribuídas à necessidade laboratorial e aos custos relativamente elevados. Mas, uma vez que os *primers* estejam disponíveis, o custo da técnica assemelha-se à de RAPD, com exceção de que os géis para resolver os fragmentos de DNA devem ser de poliacrilamida e esses são de custo mais elevado. A maior vantagem dessa técnica é o elevado polimorfismo revelado, o que a torna uma das melhores opções para uso na caracterização de cultivares, especialmente em germoplasma aparentado e de baixa variabilidade (FALEIRO *et al.*, 2001b; SERENO *et al.*, 2003; LANAUD *et al.*, 1999).

Segundo Grattapaglia (1999), dentre os marcadores existentes, os microssatélites são os marcadores geneticamente mais informativos. Eles são marcadores co-dominantes, ou seja, ambos os alelos de um indivíduo heterozigoto são visualizados no gel, bem como, multialélicos e suficientemente estáveis para serem utilizados em análises genéticas, sendo ideal para aplicação em genética e evolução de populações naturais.

2.4 Descrição do Domínio das Informações

A distância entre a geração de informação e sua utilização prática normalmente é grande em qualquer ramo da ciência. Isto é especialmente

verdadeiro no caso de utilização de marcadores moleculares no melhoramento genético de plantas. Um grande volume de informações a respeito do assunto vem sendo gerado cada vez mais rapidamente. Com o uso crescente dos marcadores, o apropriado armazenamento desses dados torna-se extremamente importante.

Neste sentido, faz-se relevante analisar o conjunto de informações a ser manipulado e armazenado em banco de dados, evitando armazenamento redundante e fornecendo formato padronizado para facilitar a recuperação, re-análise e a troca dos dados.

No caso em estudo os seguintes conceitos estão envolvidos no problema:

- Acesso, caracterizado por nome da espécie, local de origem de seu material, informações fenotípicas, como cor do fruto e informações agronômicas, como peso do fruto seco.
- Loco, caracterizado por nome do seu marcador, pelo acesso que o possui e por informações genéticas como o tamanho dos alelos presentes e o grupo de ligação ao qual se relaciona.
- *Primer*, caracterizado por nome, a seqüência que o compõe, a temperatura de anelamento*, motif* e qual tipo de marcador ele representa.
- Artigo científico, caracterizado por dados bibliográficos que possui, como título, autores, palavras-chave e outros.
- Instituição, caracterizada por nome e endereço das instituições participantes no projeto.
- Laboratório, caracterizado por nome e pela instituição que faz parte.
- Usuário, caracterizado por nome completo, dados para autenticação no sistema, dados pessoais e o laboratório em que trabalha.

2.5 Bancos de Dados de Marcadores Moleculares

Existem vários tipos de bancos de dados com diversas informações biológicas. Muitos são os que possuem dados de marcadores moleculares armazenados para fins de melhoramento genético. Alguns exemplos de bancos de dados com informações de marcadores moleculares em plantas são: Genbank, Wheat MM Database (ANWMMD), GrainGenes, Crop-Store, Arabidopsis SSR

Database (CAPSL), Plant SSR Database, International Cocoa Germplasm Database (ICGD), entre outros.

O GenBank® é um banco de dados de seqüências genéticas, uma coleção anotada de todas as seqüências de DNA públicas disponíveis. São mais de 44 bilhões de bases em mais de 40 milhões de seqüências gravadas no ano de 2004 (Estatística de crescimento do GenBank, 2005). <http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html>

O Crop-Store é um banco de dados que foi desenvolvido para combinar pesquisas de gerenciamento de recursos genéticos, mapeamento e melhoramento, e armazena informações relacionando referências de recursos genéticos, características, dados de loci de marcadores e linkage, e fontes de dados genômicos, para programas de melhoramento de diversas plantas.

O CAPSL (Arabidopsis SSR Database) é um site que remete a um banco de dados com todos os microssatélites de 20bp (pares de bases) de comprimento ou mais, com motifs de singlets para penta-nucleotídeos, encontrados no genoma de *Arabidopsis* sp.

O Plant SSR – CUGI (Plant Simple Sequence Repeat Database), busca identificar simple sequence repeats ou microssatélites em seqüências de plantas.

O GrainGenes possui serviços incluindo bancos de dados, documentos, ferramentas, arquivos de dados, sites web, avisos, entre outros, sobre aveia, cana-de-açúcar e Triticeae (subgrupo da família Gramineae que contém a maioria dos cereais: trigo (*triticum*), centeio (*secale*), e cevada (*hordeum*). O banco de dados do GrainGenes possui informações tais como: mapas genéticos e citogenéticos; sondas genômicas, seqüências de nucleotídeos; fenótipos associados; genótipos e árvores genealógicas de cultivares, e outros germoplasmas; patologias e os correspondentes patógenos, insetos, e estresses abióticos; entre outras informações importantes sobre estes cultivos.

O ANWMMMD (Australian National Wheat Molecular Marker Database) é um banco de dados que contém os resultados de uma colaboração australiana visando desenvolver microssatélites em trigo.

Enfim, existem vários tipos de bancos de dados sobre diversas plantas. Sobre a planta de estudo deste trabalho, o cacaueteiro (*Theobroma cacao* L.), existe um

banco de dados muito importante: O ICGD (International Cocoa Germplasm Database).

O projeto do ICGD, do Reino Unido, foi iniciado em resposta a uma necessidade global de serviços de informação de germoplasma para a comunidade de pesquisa do cacau; em particular, produtores de cacau e administradores de banco de genes.

Um considerável volume de informação sobre clones de cacau sempre existiu, mas nunca extensamente disponível além do instituto de pesquisa no qual ele foi coletado. Este problema é aumentado pelo fato de que as maiores áreas de crescimento estão geograficamente distantes, limitando a interação e o fluxo da informação. A meta do projeto foi confrontar esses dados em um banco de dados computadorizado e circular informação através da comunidade global de cacau por meio de software e publicações. Dessa forma o ICGD oferece uma plataforma para a transferência dos dados e conhecimentos sobre clones de cacau e pesquisa e benefícios de cacau e de trabalhadores de todos os níveis.

O banco de dados proposto neste trabalho difere do banco ICGD por possuir formatação diferenciada própria, direcionada aos usuários da região cacaueira do Sul da Bahia, além de ter uma configuração toda em português, para viabilizar maior facilidade na utilização por parte de produtores e conter dados do Brasil e, mais precisamente, da região sul-baiana.

3. METODOLOGIA

São apresentadas, neste item, as ferramentas utilizadas no desenvolvimento do Sistema de Informação, a utilização de tais ferramentas e suas funcionalidades.

3.1 MATERIAIS

3.1.1 Hardware

Para o desenvolvimento deste sistema utilizou-se de uma máquina Dell, Pentium 4, 1 GHz, 512 Mb de RAM, sob um sistema operacional Linux, (Fedora Core 2). Por sua vez, os servidores WEB e de Banco de dados que compõem a arquitetura do sistema são máquinas bi-processadas xeon, 1.7 GHz, 2 Gb de RAM e que rodam sob uma plataforma Linux (Fedora Core 2).

3.1.2 Softwares e Linguagens

Os critérios para seleção dos softwares e linguagens levaram em conta as necessidades requeridas para o desenvolvimento do Sistema e que preferencialmente fossem de distribuição gratuita e rodassem em plataforma Linux.

3.1.2.1 GNU Linux

A plataforma escolhida para manter o sistema proposto foi o GNU Linux, uma variante do UNIX, porém gratuita. Primeiramente por esta ser a atual plataforma utilizada nos servidores do laboratório em que será mantido o sistema, Laboratório de Bioinformática (LABBI / UESC).

3.1.2.2 UML

Para modelar o sistema foi utilizada a UML, Linguagem de Modelagem Unificada, principalmente por ter uma notação diagramática para desenvolvimento de sistemas de

software que fornece uma abordagem abrangente nas etapas de análise e projeto, a partir de uma perspectiva orientada a objetos.

UML será usado já que apresenta características essenciais para a criação de sistemas de software bem-projetados, robustos e de fácil manutenção, principalmente por dispor de princípios para elaboração de uma solução lógica (LARMAN, 2000).

3.1.2.3 Poseidon

Para a construção dos artefatos UML foi usado o Poseidon for UML Community Edition que é gratuito, suporta vários diagramas UML, de fácil utilização e aprendizado, além de permitir a exportação dos diagramas em formatos como JPEG e GIF.

3.1.2.4 MySQL

No desenvolvimento do banco de dados foi utilizado o SGBD MySQL em sua versão 5.0 para Linux, por ser gratuito, confiável, rápido, requerer baixa demanda de recursos de hardware, alta performance e oferecer recursos de controle de integridade referencial (GONZAGA, 2002; WELLING e THOMSON, 2001).

3.1.2.5 PHP

Para compor a lógica de aplicação do sistema a linguagem PHP foi escolhida por permitir o desenvolvimento de aplicações dinâmicas e de alta performance, para ser utilizada na Internet. Além de permitir a interação direta do usuário com o site, visto que as informações são processadas em tempo real, juntamente com as de outros usuários, retornando em páginas HTML dinâmicas geradas para cada usuário. No que se refere à segurança, é server-side, o que permite a integridade das aplicações desenvolvidas com esta linguagem, uma vez que o código é interpretado no servidor e somente o resultado deste processamento é enviado para o cliente (PHP, 2003; WELLING e THOMSON, 2001; GREENSPAN e BULGER, 2001; GONZAGA, 2002).*

3.1.2.6 JavaScript

Esta linguagem permite tornar a página HTML mais interativa e funcional. É executada pelo WEB-Browser do cliente quando a página HTML que o contém é visualizada e foi escolhida para diminuir a complexidade da lógica de programação, já que algumas verificações são antecipadas e ocorrem ainda na máquina do cliente.*

3.1.2.7 Fireworks

Este editor de imagens foi escolhido devido os seus recursos para produção, desde pequenas imagens, até grandes e complexos layouts de portais. Isso sem contar que possui uma desejável integração com o Dreamweaver e com o HTML.

3.1.2.8 Corel Suíte

Este é um pacote de programas de desenhos gráficos, edição de imagens e com suporte a HTML. Como o Fireworks, este possui recursos poderosos para o desenho de uma interface leve e agradável ao usuário, e por estas características foi selecionado.

3.1.2.9 Dreamweaver

A fim de facilitar a edição do código HTML e dos scripts PHP foi escolhido o Dreamweaver, pois auxilia na utilização dos objetos e funções da linguagem PHP permitindo a edição dos scripts de forma rápida.*

Apesar de ter licença paga, pois é uma ferramenta fornecida pela empresa Macromedia, foi utilizada na criação de páginas-protótipo do sistema uma versão trial (de 30 dias), até que seja decidido qual editor de scripts de licença livre, que se aplique em SO Linux, será empregado na elaboração da interface WEB do sistema.

3.1.2.10 Programa GENES

O Programa GENES é um software distribuído gratuitamente destinado à análise e ao processamento de dados por meio de diferentes modelos biométricos. Visa auxiliar o melhorista nas tomadas de decisões e escolha de estratégias que permitam aumentar a chance de êxito de um programa de melhoramento genético, por meio do processamento de dados experimentais.

O uso do Programa GENES é de grande importância em pesquisas genéticas aplicadas ao melhoramento vegetal e animal, por permitir estimar parâmetros estatísticos para entendimento de fenômenos biológicos e fundamentais em processo de tomada de decisão e na predição do sucesso e da viabilidade da estratégia de seleção, além de produzir informações úteis no monitoramento e na conservação da biodiversidade (CRUZ, 1997).

Por causa da importância desse aplicativo no estudo de diversidade genética, o sistema estará estruturado para gerar arquivos com a finalidade de servir de entrada para ser analisado por ele. O Programa GENES reconhece arquivos-texto e arquivos-planilha e

exige que estes estejam num formato próprio. Portanto, para que o Sistema seja capaz de produzir tais arquivos é necessário estudar os padrões utilizados. Depois de definidos os padrões, programa-se o Sistema para que este possa captar os dados do BD e reverter o arquivo que os contém no tipo de arquivo a ser usado.

3.2 MÉTODOS

3.2.1 Modelo de Desenvolvimento

Para o desenvolvimento do Sistema de Informação para Apoio ao Programa de Melhoramento Genético do Cacaueiro, foi adotado um modelo iterativo de tipo incremental por camadas. Esta escolha proporcionou maior flexibilidade no planejamento do projeto e permitirá a entrega regular de versões funcionais do sistema cada vez mais aperfeiçoadas e ampliadas.

Em cada iteração executar-se-á uma seqüência padrão de atividades, ou workflows, consistindo de levantamento de requisitos, análise, projeto, implementação, testes e liberação.

Na primeira iteração, foi abordada a camada de armazenamento dada a prioridade de construir primeiramente um banco de dados relacional, e na segunda (em andamento) foi considerada a camada de interface. A camada de lógica de aplicação será objeto da próxima iteração. A seguir descrevem-se os métodos utilizados para as diversas fases das duas primeiras iterações.

3.2.2 Levantamento de Requisitos.

Como ponto de partida para o levantamento de requisitos, foi consultado o projeto de pesquisa “Estabelecimento de sistema automático e informatizado para análises de diversidade genética e genotipagem de acessos de cacaueiro Theobroma cacao L. com base em marcadores moleculares microssatélites” (CORRÊA et al., 2002), no qual este trabalho se insere. Neste documento foram identificados os objetivos gerais e específicos do sistema.

Posteriormente foram entrevistados seis pesquisadores: dois participantes do projeto no qual este trabalho se insere, dois envolvidos no programa de melhoramento genético do cacaueiro da CEPLAC e dois com conhecimentos gerais na área. Três produtores da região foram também consultados para este fim.

Visando uma melhor estruturação das informações obtidas, foi elaborado um formulário semi-estruturado com questões direcionadas ao levantamento de requisitos do banco de dados e funcionais e aplicado às mesmas pessoas anteriormente entrevistadas.

As informações coletadas podem ser resumidas da seguinte forma:

- *O sistema deve conter um banco de dados que combine informações genotípicas, fenotípicas e bibliográficas (artigos científicos) dos acessos de cacauero da região;*
- *O banco de dados deve abarcar informações de distintas origens por diferentes usuários;*
- *Os dados devem ser acessíveis pela Internet, podendo ser modificados apenas pelos próprios depositários, mas liberados para consultas para qualquer usuário cadastrado. A política de acesso será definida pelo coordenador do projeto;*
- *O sistema deve fazer a análise de diversidade genética e fingerprint;*
- *A interface deve ser amigável e auto-explicativa.*

Com base nessas informações foram construídos o diagrama de contexto, o modelo de casos de uso, o glossário e o protótipo da interface, e identificados as entidades e os relacionamentos do DER, bem como as classes e seus atributos, para o projeto físico do BD.

3.2.3 Análise

A análise foi baseada no paradigma orientado a objeto. Foram identificados as entidades e os relacionamentos entre elas a partir da análise sintática da descrição formal dos requisitos em duas etapas: (1) seleção dos substantivos como candidatos a entidades e os verbos a relacionamentos. (2) análise individual de cada candidato enquanto sua coerência contextual, utilizando-se argumentos lógicos do raciocínio comum.

A seguir foi realizada a identificação dos objetos relevantes ao modelo e sua posterior classificação.

Com estas informações foram construídos os diagramas de Classes, utilizando-se UML, e de Entidades-Relacionamentos⁴.

3.2.4 Projeto

Nesta fase efetuou-se a escolha da plataforma computacional e das tecnologias de desenvolvimento, a definição da estrutura do sistema em nível de componentes e definiu-se o projeto físico do banco de dados. A cada entidade do DER foi atribuída uma tabela no

⁴ O Diagrama de Entidade-Relacionamento (DER) é um resultado da análise estruturada moderna, mas foi utilizado para melhor compreensão pela autora e uma migração mais coerente para a análise baseada no paradigma orientado a objeto.

banco de dados, enquanto os relacionamentos, segundo sua cardinalidade⁵ foram transformados em chaves de acesso ou em tabelas de relacionamento (secundárias).

Os diagramas de colaboração e de instalação/distribuição não foram ainda elaborados, ficando para a próxima iteração.

3.2.5 Implementação

Utilizando-se MySQL 5.0 para GNU Linux foi implementado o banco de dados do sistema fazendo um mapeamento fiel do projeto físico.

A codificação da interface está em andamento e para tal estão sendo usados o editor de HTML e scripts Dreamweaver.

A codificação da lógica, a qual estabelecerá o acoplamento da interface com o banco, está programada para a próxima iteração.

3.2.6 Testes

Os testes de funcionalidades e integridade do banco de dados encontram-se em andamento atualmente. Para isto as tabelas estão sendo povoadas com dados provenientes do projeto “Diversidade genética de cacauero (Theobroma cacao L.) resistente à vassoura-de-bruxa com base em marcadores RAPD e microssatélites” (LEAL, 2004), e efetuadas consultas que ativam sistematicamente os diferentes relacionamentos.

⁵ A cardinalidade é um número que expressa a frequência (número de ocorrências) de determinada entidade associada a uma ocorrência da entidade em questão através do relacionamento; expressa o número de entidades às quais uma entidade pode ser associada via um conjunto de relacionamentos.

4. RESULTADOS E DISCUSSÃO

Serão apresentados neste capítulo, os artefatos resultantes das duas primeiras iterações (a segunda ainda incompleta) do desenvolvimento do Sistema de Informação para Apoio a Programas de Melhoramento Genético, consistindo basicamente de um banco de dados relacional e um protótipo de interface WEB.

4.1 Levantamento de Requisitos

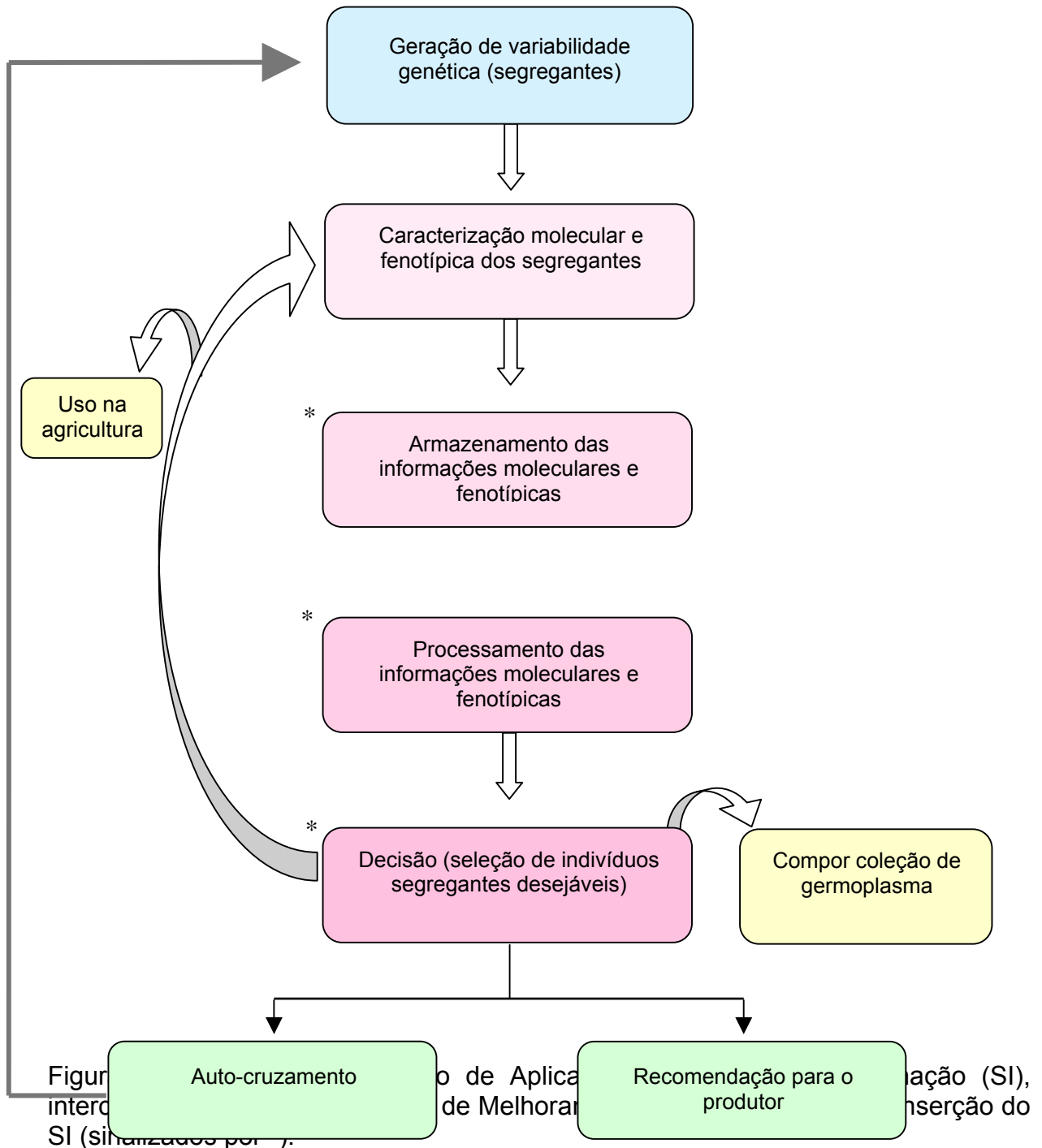
O principal objetivo desta fase foi atingido, consistindo na descrição do domínio de aplicação do sistema de informação, sob a ótica de ser um elemento integrante ativo ou passivo⁶ dentro do processo que nele acontece. Desde o ponto de vista conceitual esta é a fase de maior contribuição intelectual à solução do problema abordado, na qual se identifica a complexidade inerente ao mesmo e se projeta a transferência da mesma para o sistema a ser desenvolvido. Uma prova desta transferência de complexidade é o fato que quanto maior for a complexidade do problema abordado mais complexo será o software que resolverá ou facilitará a solução desse problema. É como se existisse uma lei de conservação da complexidade: a complexidade, uma vez criada, não pode ser destruída, mas transformada.

Pode-se dizer que o analista de sistemas é o profissional responsável por identificar o problema, capturar sua complexidade e encapsulá-la na forma de sistema, projetado para facilitar o processo associado ao problema. No caso em estudo, o problema inerente ao processo de pesquisa de apoio aos programas de melhoramento genético, está na dificuldade da extração de informação relevante (útil para a finalidade do estudo) num universo crescente de dados heterogêneos e dispersos.

⁶ Por exemplo, um sistema de informação pode ser utilizado, independentemente de suas funcionalidades, apenas para armazenar e reproduzir os dados armazenados, i.e, como uma imagem eletrônica de um armário convencional utilizado para arquivar informações, ou pode também ser explorado para buscar e processar digitalmente as informações nele armazenadas.

4.1.1 Diagrama de Contexto

O primeiro artefato construído foi o Diagrama de Contexto, que representa simbolicamente o processo atual dos programas de melhoramento genético. Este diagrama foi extremamente importante para obter uma visão geral do domínio do problema e estabelecer os pontos onde o sistema de informação poderia ser inserido (Figura 6).



4.1.2 Modelo de Casos de Uso

Com base no Diagrama de Contexto e nos requisitos levantados a partir das entrevistas realizadas, foram identificados diversos cenários de utilização do sistema por parte dos usuários, i.e, os Casos de Uso. Observou-se que este processo não foi conclusivo e que no decorrer do projeto alguns Casos de Uso foram incluídos ou modificados e outros foram excluídos. No Quadro 3 abaixo segue a lista atual dos Casos de Uso e as respectivas finalidades. Os detalhes de cada Caso de Uso, incluindo o fluxo principal e os sub-fluxos, podem ser consultados no Apêndice 1.

Quadro 3. Casos de Uso e suas respectivas funcionalidades.

Nº	Casos de Uso	Finalidade
1	<i>Manipular Dados de Acesso</i>	<i>Atualizar os dados dos Acessos de Cacaueiro no BD</i>
2	<i>Manipular Dados de Loco</i>	<i>Atualizar os dados dos Locos no BD</i>
3	<i>Manipular Dados de Primer</i>	<i>Atualizar os dados dos Primers no BD</i>
4	<i>Ativar Formulário</i>	<i>Disponibilizar para o usuário os serviços de Inserção, Deleção, Atualização e Impressão de informações do BD, comuns a todos os formulários de interface</i>
5	<i>Cadastrar Artigo</i>	<i>Cadastrar Artigos Científicos no BD</i>
7	<i>Cadastrar Autor</i>	<i>Cadastrar Autores de Artigos Científicos no BD</i>
8	<i>Realizar Consultas Avançadas</i>	<i>Criar relatórios a partir de consultas personalizadas</i>
9	<i>Pesquisar Artigos</i>	<i>Pesquisar o material bibliográfico existente no BD</i>
10	<i>Cadastrar Usuário</i>	<i>Cadastrar um visitante da página como usuário do BD</i>
11	<i>Autenticar Usuário</i>	<i>Permitir o acesso ao Sistema de Informação aos usuários cadastrados</i>
12	<i>Cadastrar Instituição</i>	<i>Cadastrar Instituição de pesquisa da qual o usuário faz parte</i>
13	<i>Cadastrar Laboratório</i>	<i>Cadastrar laboratório de trabalho do usuário</i>

14	Criar grupo de trabalho	Criar um grupo de trabalho formado por usuários
----	-------------------------	---

Baseado na especificação formal de requisitos, foram também identificados os diferentes tipos de usuários ou atores e sua relação com os Casos de Uso. O Diagrama de Casos de Uso mostrado na Figura 7 ilustra este aspecto.

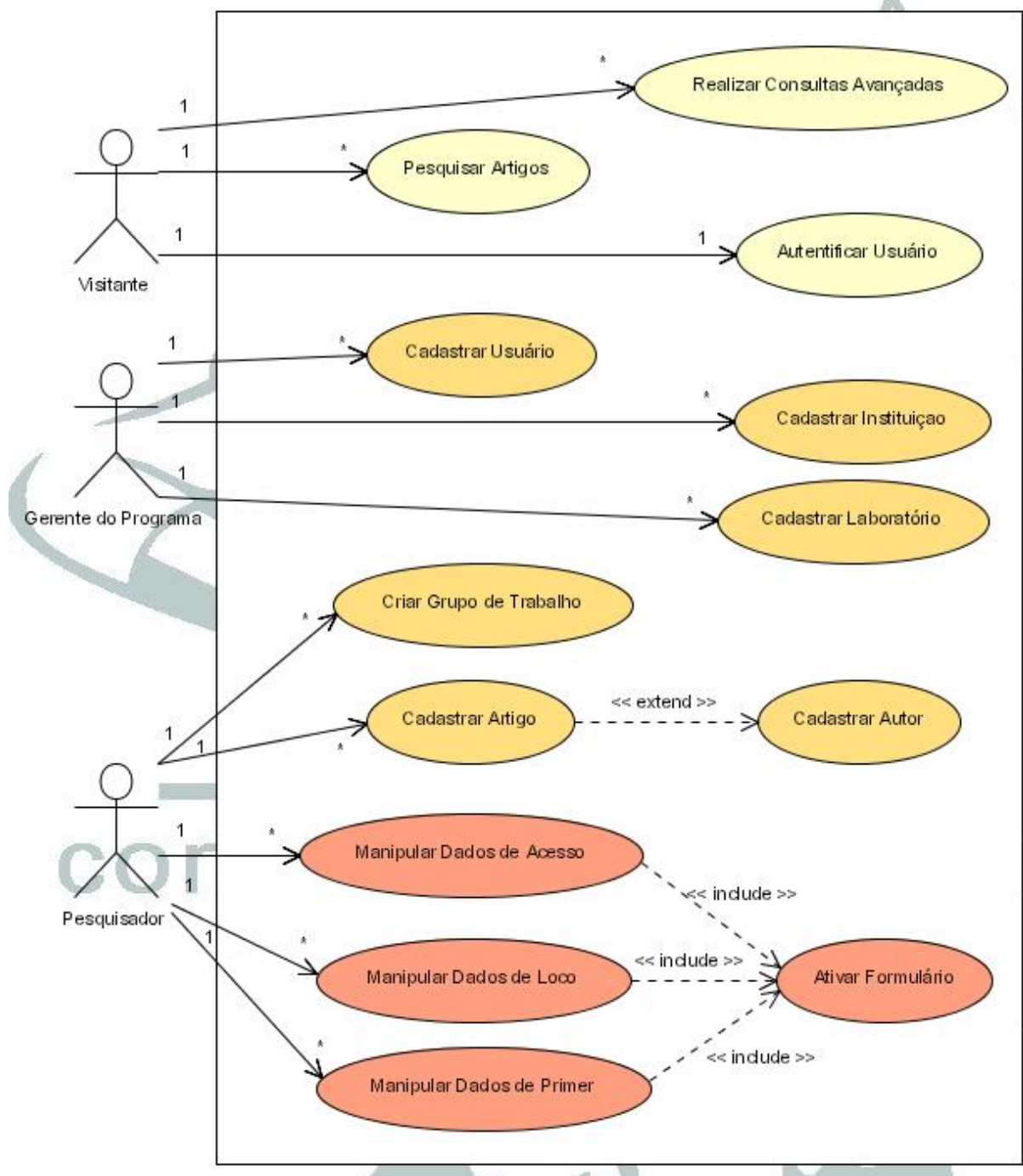


Figura 7. Diagrama de Casos de Uso e sua alocação nos *releases* seguindo o padrão de uma cor por vez.

Neste artefato podem ser observados os atores⁷ do sistema, representados por bonecos, os casos de uso, pelas formas ovais, e a cardinalidade das relações entre eles, representados pelos símbolos (1 e *) sobre as setas. Por exemplo, a relação entre o “Ator Usuário” e o Caso de Uso “Cadastrar Artigo” é de um para muitos, ou seja, um usuário pode cadastrar zero ou muitos artigos. As setas contínuas, seguindo a sua orientação, representam uma ação e qual ator a inicia e as tracejadas indicam inclusão (uso obrigatório) ou extensão (uso opcional) dos casos de uso relacionados. O limite físico do Sistema de Informação é representado pelo retângulo que marca quais funcionalidades o sistema possuirá.

Neste Diagrama de Casos de Uso o ator “Visitante” é qualquer indivíduo que queira pesquisar sobre qualquer tipo de informação que se encontrar disponível no sistema. O ator “Gerente de projeto” é o responsável (ou grupo de pessoas responsáveis) pelo cadastramento do ator “Visitante” afim de que este torne-se um outro tipo de ator: o ator “Pesquisador”. Além disso, o “Gerente de projeto” também cadastra Instituições e Laboratórios, e se responsabiliza em gerenciar os dados depositados pelos “ Pesquisadores”, tornando-os acessíveis ou não para os “Visitantes”, dependendo da disponibilização por parte dos “Pesquisadores”. O ator “Pesquisador” é aquele indivíduo que pode inserir, manipular ou excluir seus dados no sistema, além de decidir por disponibilizar ou não para outros “Pesquisadores”, ou até mesmo para os “Usuários”, dados próprios que ele julgue disponibilizáveis no sistema.

Foi considerado que o sistema será de caráter público, contendo dados públicos e dados de acesso restrito, de acordo com as permissões dos “Pesquisadores” proprietários dos dados em questão. Este sistema permitirá o depósito de dados mediante cadastro de usuário pelo gerente do projeto, e os dados serão sempre referenciados ao usuário depositário.

Os casos de uso são agrupados por cores de acordo com sua prioridade. Cada *release* consta da implementação e liberação de um grupo por vez, a começar pela vermelha, tida como a atividade mais crítica, seguida do laranja e amarelo.

⁷ É a entidade que inicia ou participa de um caso de uso, sendo que este ator pode ser um usuário de sistema, um caso de uso ou um outro sistema.

4.1.3 Glossário

O glossário (ou dicionário de dados do modelo) foi implementado para melhorar a comunicação entre os membros do projeto e reduzir os riscos de mal-entendidos, pois lista e define todos os termos que requerem esclarecimentos. Assim sendo, tais informações encontram-se na seção Glossário neste trabalho.

4.1.4 Protótipo da interface

Embora este segmento do projeto não tenha relação direta com a funcionalidade lógica do sistema, ele se relaciona com o grau de aceitação do produto pelo usuário. E, por ser um requisito do sistema, foi desenvolvido primando-se pela facilidade de utilização, objetividade, simplicidade e beleza para satisfazer as necessidades do usuário final.

Como o Sistema proposto tem como uma de suas finalidades automatizar o trabalho de pesquisadores, transferindo a complexidade do mesmo para o processamento interno do sistema, boa parte do sucesso deste fim é relacionada ao bom projeto da interface, único componente do sistema em contato direto com o usuário. Assim sendo, a interface possui a importante missão de capturar as intenções do usuário, solicitar processamento ao sistema e apresentar os resultados encontrados de maneira simples, objetiva e de fácil entendimento. São amostras das interfaces prototipadas as Figuras 8 e 9.

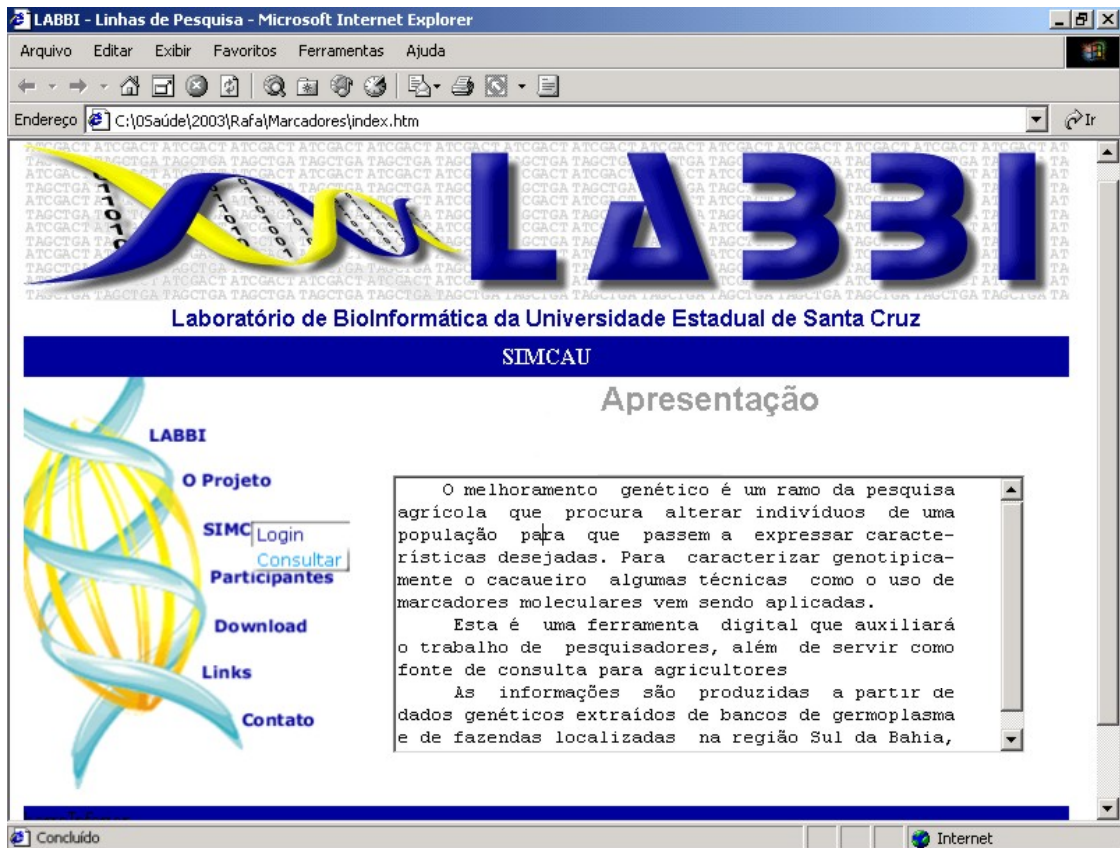


Figura 8. Página principal da Interface.

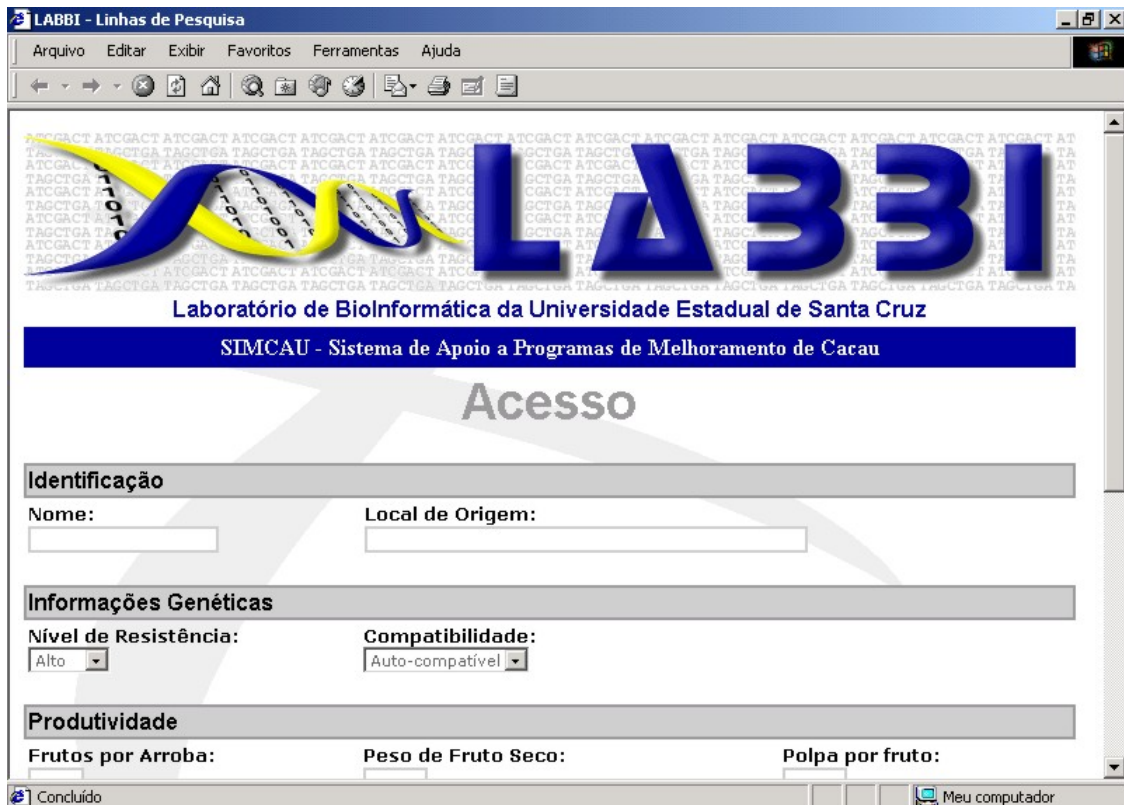


Figura 9. Formulário de interface para manipulação de informações de acessos de cacauero.

4.2 Análise

4.2.1 Diagrama de Entidades e Relacionamentos (DER)

O DER mostrado na Figura 10 foi o primeiro artefato ligado à implementação do Banco de Dados Relacional a ser produzido, pois teve o propósito de promover a discussão sobre quais entidades seriam relevantes para a definição da estrutura do banco de dados.

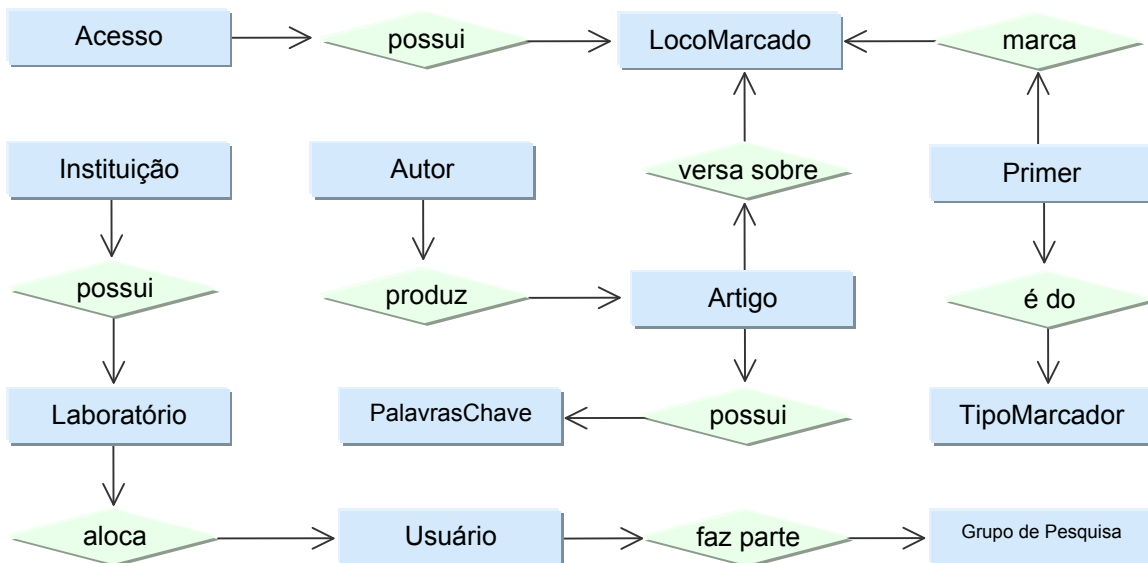


Figura 10. Diagrama Entidade-Relacionamento simplificado (sem a cardinalidade), no qual as entidades são indicadas por retângulos e os relacionamentos, por losangos.

Além das entidades (como Acesso, Loco e Primer) representadas pelos retângulos, o DER também apresenta as relações entre elas (como possui, marca e fala de) representadas pelos losangos. Por exemplo, Acesso possui Loco.

Neste artefato pode-se observar, por exemplo, a relação existente entre 'Primer' e 'Loco', em que o primer *marca* o loco; ou entre 'Primer' e 'TipoMarcador': em que o primer *é de um* tipo de marcador.

Com a definição das entidades e dos seus relacionamentos foi possível realizar uma análise que correspondesse propriamente ao paradigma orientado a objeto.

4.2.2 Diagrama de Classes

O Diagrama de Classes foi criado para enfatizar quais entidades de *software* serão necessárias para a solução lógica. Esta etapa marca o início da transformação dos conceitos do mundo real em entidades de *software*.

Por meio da Figura 11, são mostradas as classes e suas associações. Cada classe tem definido os seus atributos (idNome, origem, corFruto), por exemplo e operações (insere(), busca(), analisa()). As associações simbolizam relações entre as classes, indicando a multiplicidade (número de objetos relacionados) entre elas, onde o símbolo 1 representa que apenas uma unidade do objeto faz parte da relação e o símbolo * representa que zero ou muitos objetos participam dela. Por exemplo, de acordo com o diagrama de classes abaixo, um *Acesso* possui zero ou muitos *Locos*.

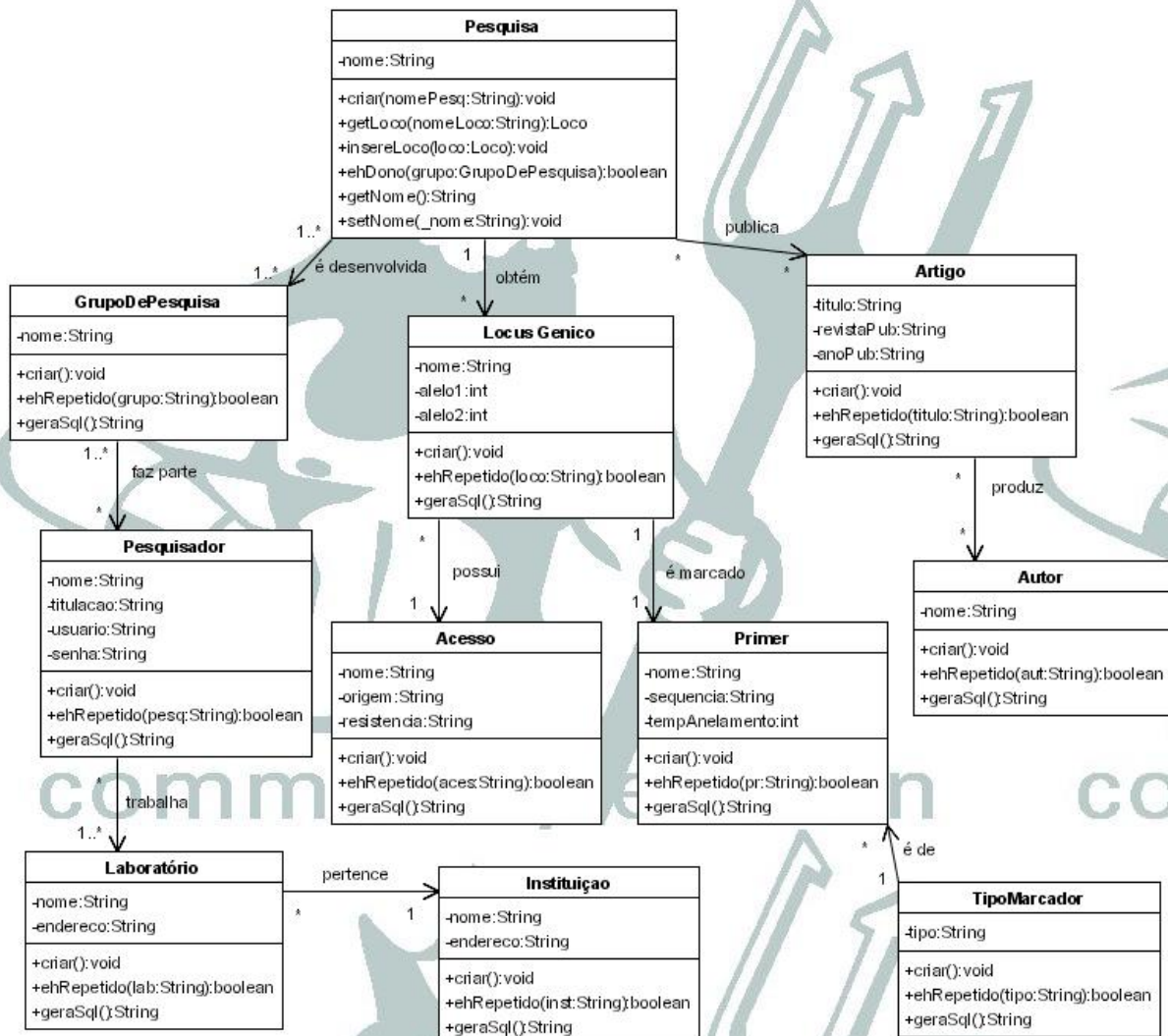


Figura 11. Diagrama de Classes.

4.3 Projetos

4.3.1 Diagrama de Componentes

A Figura 12 mostra a arquitetura definida para o sistema, que foi organizado de modo a apresentar uma interface com o usuário, a aplicação e uma forma de armazenamento consistente de dados (o banco de dados).

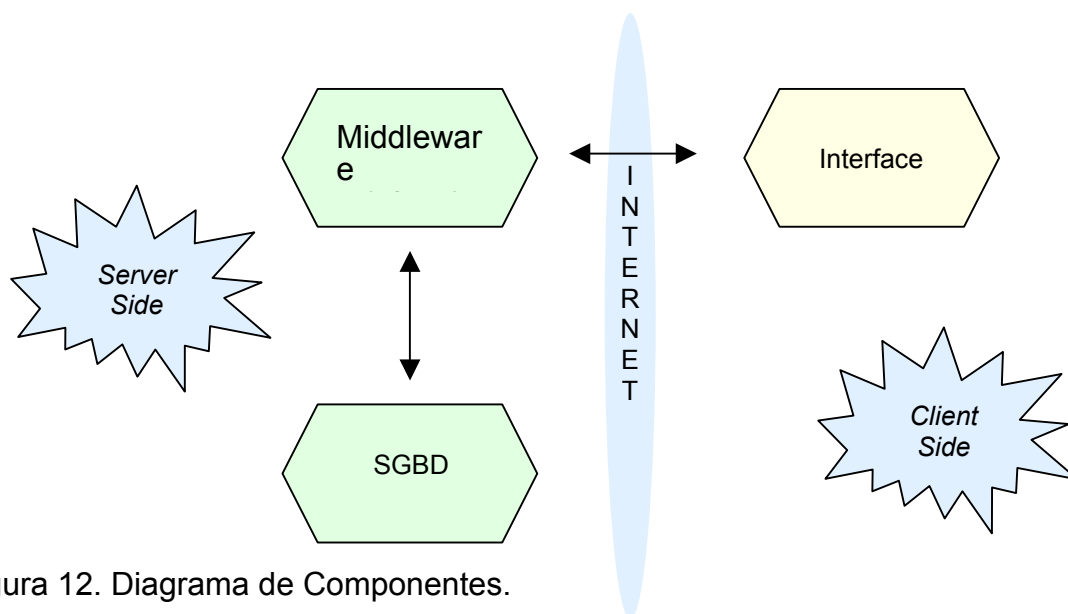


Figura 12. Diagrama de Componentes.

Observa-se também que a aplicação e as consultas ao Banco de Dados são processadas do lado do servidor (*server side*), de forma a preservar a segurança da aplicação, uma vez que o usuário não tem acesso ao código-fonte do sistema e exigir menos recursos de máquina do usuário já que eles recebem páginas dinâmicas em HTML geradas pela aplicação de acordo com as solicitações que enviam via interface WEB.

4.3.2 Projeto do Banco de Dados

A partir das entrevistas e da aplicação do formulário semi-estruturado (vide anexo) foi possível a elaboração deste artefato que define a estrutura do banco de dados e serve de consulta para a codificação em SQL (Figura 13). Os detalhes de cada tabela do projeto do BD podem ser consultados no Apêndice 2.

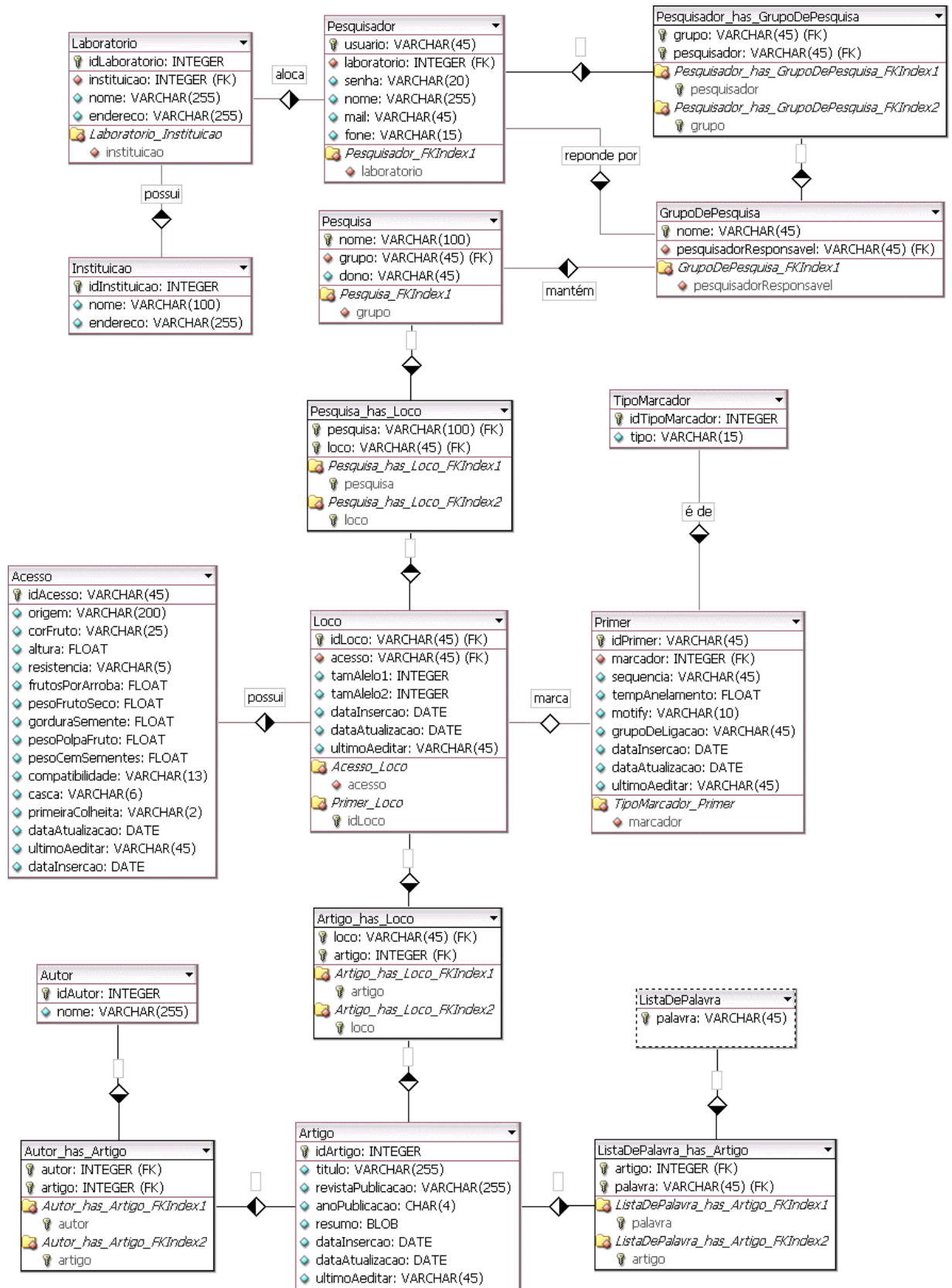


Figura 13. Projeto físico do banco de dados relacional, mostrando as tabelas e seus relacionamentos.

4.4 Implementação

4.4.1 Banco de Dados MySQL

O BD foi criado por um conjunto de comandos (*create*, *alter*, *drop*) que possibilitaram criar, alterar e deletar tabelas, e definir suas estruturas, as colunas (campos ou atributos), além das chaves primárias, as chaves estrangeiras relacionadas a elas e demais restrições de integridade. A Figura 14 mostra parte dos códigos utilizados para a construção do BD.

```
CREATE TABLE Acesso (  
  idAcesso VARCHAR(45) NOT NULL,  
  origem VARCHAR(200) NOT NULL,  
  corFruto VARCHAR(25) NOT NULL,  
  altura FLOAT NULL,  
  resistencia VARCHAR(5) NOT NULL,  
  .  
  .  
  dataInsercao DATE NOT NULL,  
  dataAtualizacao DATE NOT NULL,  
  ultimoAeditar VARCHAR(45) NOT NULL,  
  PRIMARY KEY(idAcesso)  
)  
TYPE=InnoDB;
```

Figura 14. Representação de códigos para a construção do BD, utilizando SGBD MySQL.

4.4.2 Manipulação de dados do Banco de Dados

O BD pode ser manipulado com um outro grupo de comandos básicos, como *select*, *insert*, *update* *delete*, responsáveis pela consulta e atualização dos dados armazenados em um BD. Assim, observa-se no exemplo abaixo a sintaxe utilizada para inserção das informações de um *primer* no BD do projeto:

```
INSERT INTO  
<primer> ([<nome>],  
[<sequencia>],  
[<tempAnelamento>],
```

Figura 15. Comando para inserção de dados no BD.

Agora, imaginando que o usuário quisesse solicitar a seguinte informação à interface do BD: “Qual a temperatura de anelamento do *primer* Y 16985?”. A lógica de programação, que é representada pelo PHP, geraria uma consulta ao banco de dados que satisfizesse a solicitação do usuário :

```
SELECT
```

Figura 16. Comando para consultar informação do BD.

Nesta sintaxe o comando requer que seja selecionada a temperatura de anelamento da tabela *Primer*, onde seu *idPrimer* é de valor igual a Y 16985.

4.5 Testes

O Sistema em desenvolvimento está em fase de teste α do BD. Está sendo validada a codificação e o próprio sistema faz os testes de integridade referencial* entre as tabelas.

É importante ressaltar que o desenvolvimento deste sistema de informação tem grande impacto no que diz respeito à transferência de complexidade do contexto do problema, visto que pesquisadores possuíam muitas informações dispersas e em quantidade, e o sistema se propõe organizar e personalizar consultas que se queira fazer sobre algo pontual. O desenvolvimento deste sistema baseia-se num projeto com característica inter-institucional (UESC-CEPLAC/CEPEC), em que há a necessidade de compartilhamento de dados e resultados, de forma a melhor conduzir estudos para os programas de melhoramento, partindo de informações relativas de modo que o sistema tivesse como propriedade ser capaz de apoiar a

tomada de decisões por parte de pesquisadores e, depois das outras iterações, auxílio nas decisões de produtores.

O uso de dados reais gerados no projeto de pesquisa conferem credibilidade aos testes do sistema, permitindo a avaliação do cumprimento dos requisitos relacionados com o armazenamento das informações genéticas.

5. CONCLUSÕES

A criação de um sistema de informação para apoio a programas de melhoramento genético mostra-se uma alternativa coerente, pois visa facilitar o trabalho do pesquisador nestes programas, transferindo parte da complexidade do problema para a aplicação (sistema). Além disso, provê interação entre pesquisadores e produtores, visando uma maior troca de dados entre todos os envolvidos em programas de melhoramento genético: grupos de pesquisa e grupo de usuários das tecnologias criadas para a prática agrícola.

A fim de contemplar os objetivos traçados para a realização deste trabalho, foi feito o projeto de sistema para uso específico em programas de melhoramento genético de cacau (*Theobroma cacao* L.), a partir de um estudo aprofundado referente ao processamento de informações.

Neste estudo, foram identificados os dados, as interfaces e os usuários do sistema, que permitiu desenhar um banco de dados, as interfaces e esboçar a lógica funcional necessária.

A partir do projeto foi construído um banco de dados relacional que está sendo povoado com informações de plantas de interesse regional de marcadores moleculares provenientes do projeto “Diversidade genética de cacauero (*Theobroma cacao* L.) resistente à vassoura-de-bruxa com base em marcadores RAPD e microssatélites” (LEAL, 2004).

O projeto veiculado para a construção de uma página WEB, levando em conta que diferentes usuários possuem diferentes necessidades de informação para auxiliar em sua tomada de decisão, justifica a implementação deste sistema disponibilizado em diferentes níveis de abordagens, para pesquisador e para produtor. Para isto, foram definidas a estrutura e as funções da página na WEB. Porém, devido a preocupação em fornecer ao produtor informações mais acessíveis

e que utilizem linguagem apropriada para melhor compreensão, a implementação da interface WEB para este grupo de usuários será o resultado da segunda iteração no desenvolvimento do sistema.

A estrutura e as características do banco de dados de marcadores genéticos, assim como dos procedimentos para sua manipulação foi de difícil definição, devido à insipiência com a prática de levantamento de requisitos e análise, a fim de elaborar um melhor projeto de desenvolvimento do sistema somado à dificuldade inerente a própria atividade. Contudo, este fato não afeta a capacidade e a performance do sistema, agora em sua 1ª iteração, de contemplar os requisitos de maneira funcional.

O sistema gerado é de extrema relevância e de grande proveito para as instituições de pesquisa e agricultores, colaborando para melhor utilização das informações e conhecimentos gerados nos diversos projetos de pesquisa relacionados à área. Também é importante por ampliar as relações entre pesquisadores, bem como entre produtores e pesquisadores, aperfeiçoando-se os programas de melhoramento de um cultivo de importância sócio-econômica, como o cacau.

Além disso, a relevância desse sistema se acentua pelo contexto em que se insere, pois é um sistema pioneiro no Brasil, de fundamental importância, sobretudo regionalmente já que será povoado com informações extraídas principalmente de cultivares da região cacaeira do Sul da Bahia, no entanto não se restringirá a limitações geográficas por absorver a funcionalidade de distribuição da tecnologia WEB. Semelhante a este sistema, como já mencionado no item relativo aos 'Bancos de Dados de Marcadores Moleculares', existe o ICGD – International Cocoa Genome Database - do Reino Unido, mas que, no entanto, não contempla de modo pleno as necessidades dos pesquisadores e agricultores locais, pois não trata de dados regionais e por outros motivos expostos.

O Banco de Dados foi projetado para abarcar informações provenientes das pesquisas atualmente desenvolvidas, mas observando-se a possibilidade de futuras

expansões de forma a moldar-se à realidade em que se encontra e melhor atender às necessidades, o banco de dados está desenhado para suportar a escalabilidade*, isto é, admitir sua ampliação na medida em que forem inseridas informações transcriptômica ou proteômica, por exemplo.

Uma das perspectivas de curto a médio prazo é implementar um módulo de consultas de dados, no qual o usuário do sistema possa configurar suas próprias consultas, reduzindo a dependência dele às consultas pré-estabelecidas pelo projetista do sistema e que limitam a pesquisa que este usuário deseja realizar. Assim o sistema terá características semelhantes às de um *datawarehouse**.

Ainda visando a melhoria do funcionamento do sistema, tem-se a perspectiva a longo prazo de adicionar um módulo de Apoio à Tomada de Decisão, o qual seria formado por: (1) um modelo de domínio baseado no estudo e compreensão dos problemas da aplicação vivenciados em decisões anteriores tomadas nos programas de melhoramento e (2) algoritmos de Inteligência Artificial capazes de confrontar um novo caso a ser analisado com o modelo de domínio, recuperando o caso mais próximo do problema apresentado, adaptando-o a uma solução adequada, a qual seria indicada ao usuário e incorporando o novo caso gerado ao modelo de domínio.

O projeto de desenvolvimento deste sistema prevê a construção da página implementando o acesso ao banco de dados de marcadores moleculares por parte dos pesquisadores e disponibilizando os resultados formatados para softwares aplicativos de análise de diversidade genética, sendo que o aplicativo utilizado como padrão nas primeiras iterações é o Programa GENES, de forma acessível para consulta e interação dos pesquisadores.

Além disso, pretende-se que o sistema permita uso de softwares existentes na comunidade científica e outros, para análises genéticas, posteriormente desenvolvidos neste projeto com finalidade de utilização on-line por pesquisadores e outros interessados.

Observa-se, por fim, que um projeto de segurança de informações deve ser elaborado, uma vez que informações confidenciais estarão em trânsito na Web. Tal problema poderia ser resolvido com a utilização de um protocolo de comunicação seguro como o HTTPS, por exemplo, ou com o emprego de outras técnicas de segurança.

6. REFERÊNCIAS BIBLIOGRÁFICAS

BORÉM, A. Melhoria de espécies cultivadas : Viçosa, Universidade Federal de Viçosa, v.1, p.51-94, 1999.

CEPLAC. Clonagem. Disponível em: <http://www.ceplac.gov.br/clones.htm> Acesso em maio de 2004

CORRÊA R. X.; FALEIRO, F. G. e AHNERT, D. Estabelecimento de sistema automático e informatizado para análises de diversidade genética e genotipagem de acessos de cacauero *Theobroma cacao* L. com base em marcadores moleculares microssatélites. Projeto de Pesquisa, 017/02. Câmara de Pesquisa e Pós-Graduação. Ref. CI 113, Universidade Estadual de Santa Cruz, 2002

CRUZ, C. D. e SCHUSTER, I. Programa GQMOL: *Genética Quantitativa e Molecular*, Viçosa: UFV, 2001

CRUZ, C. D. e VIANA, J. M. S. A methodology of genetic divergence analysis based on sample unit projection on two dimensional space. *Revista Brasileira de Genética*. 17, 69-73, 1994

CRUZ, C. D. Programa Genes: aplicativo computacional em genética e estatística. Viçosa: UFV, 442p. 1997

DATE, C. J. Introdução a Sistemas de Bancos de Dados. Editora Campus, 2001

Documentação do PostgreSQL 7.4.1, Disponível em: www.javalinux.com.br/pg74/preface.html Acesso em: 17/01/2005

EXCOFFIER, L.; SMOUSE, P.E. and QUATTRO, J. M. AMOVA - Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131, 479–491, 1992

FALEIRO, F. G.; LOPES, U. V.; YAMADA, M. M.; GRAMACHO, K.; PIRES, J. L.; BAHIA, R. C. S.; SANTOS, R. C.; GOMES, L. M. C.; ARAÚJO, I. S.; FALEIRO, A. S. G.; MELO, G. R. P.; MONTEIRO, W. R.; VALLE, R. R. Caracterização Molecular de Variedades Clonais de *Theobroma cacao* L. Recomendadas pelo CEPEC/CEPLAC com Base em Marcadores RAPD, Microsatélites e AFLP. CEPEC/CEPLAC. Resumo, Ilhéus, BA, 2001a

FALEIRO, A. S. G.; FALEIRO, F. G.; LOPES, U. V.; MELO, G. R. P.; YAMADA, M. M.; BAHIA, R. C. S.; CORRÊA, R. X. Diversidade Genética de Acessos de *Theobroma cacao* L. selecionados por produtores para resistência à vassoura-de-bruxa com base em marcadores microsatélites. XXXV Congresso Brasileiro de Melhoramento de Plantas. Goiânia, 2001b

FERREIRA M. E. e GRATTAPAGLIA D. Introdução ao uso de marcadores moleculares em Análise Genética. Embrapa. Brasília, DF, 1998

FRISCH, M.; LAMKEY, K.R.; MELCHINGER, A. E. Storage of Molecular Marker Data in Databases for Efficient Use in Plant Breeding Programs. *Zeitschrift für Agrarformatik*, 10, 23-27, 2002

GIBAS, C. e JAMBECK, P. Desenvolvendo Bioinformática. Editora O' Reilly, 464p, 2002

GREENSPAN, J. e BULGER, B. MySQL / PHP Database Applications. M&T Books, 2001

GRIFFITHS, A. J. F *et all*. Genética Moderna. Ed. Guanabara Koogan, 1ª ed., 589p, 2001

KOSHAFIAN, S. Banco de Dados Orientado a Objetos. Infobook, Rio de Janeiro, 353p, 1994

LARMAN, C. Utilizando UML e Padrões: Uma introdução à análise e ao projeto orientados a objetos. Porto Alegre, RS. Editora Bookman, 608p, 1999

LEAL, J. B. Diversidade genética de cacauzeiros (*Theobroma cacao* L.) resistentes à vassoura-de-bruxa com base em marcadores RAPD e microssatélites. Dissertação de mestrado. Universidade Estadual de Santa Cruz – UESC, Ilhéus, 58p, 2004

LEWIS, P.O. e ZAYKIN, D. Genetic Data Analysis: Computer Program for the Analysis of Allelic Data, versão 1.0, 1997

LEWIS, P.O. e ZAYKIN, D. Genetic Data Analysis: Computer Program for the Analysis of Allelic Data, versão 1.0 for Windows 3.1 1999

LINGUAGENS DE PROGRAMAÇÃO Disponível em:
http://professores.faccat.br/azambuja/index_paradigmas.htm, Acesso em: maio de 2004

MAFRA, C.L.; PAULA, S. O.; OSAKI, L.S.; PATARROYO, J.H. Useful WEB sites for researchers studying proteins. *Brazilian Journal of Medical and Biological Research*, 30, 283-288, 1997

PAULA Filho, W. P. Engenharia de Software: Fundamentos, Métodos e Padrões. LTC Editora, 2ª Ed., Rio de Janeiro, 602p, 2003

PEREIRA, S.L., RAM, A., FIGUEIREDO, I.M., ALMEIDA, L.C.C. Primeira ocorrência da Vassoura de Bruxa na principal região produtora de cacau do Brasil. *Agrotrópica*, 1. 79-81, 1989

PINTO, L. R. M. e PIRES, J. L. Seleção de plantas de cacau resistentes a vassoura-de-bruxa. Ilhéus. CEPLAC/CEPEC. *Boletim Técnico*, 181, 1-35, 1998

PROSDOCIMI, F. *et al.* Bioinformática: Manual do Usuário. *Biotecnologia, Ciência & Desenvolvimento*, 29, 12-25, 2002

SCHNEIDER, S. ROESSLI, D and EXCOFFIER, L. Arlequin ver. 2000: A Software for Population Genetics Data Analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland, 2000

SILVA Filho, A. M. da. Arquitetura de Software - Desenvolvimento Baseado na Arquitetura, Editora Campus, 240p, 2002

STAIR, R. M. Princípios de Sistemas de Informação – Uma Abordagem gerencial. Rio de Janeiro, LTC – Livros Técnicos e Científicos, 1998.

TANENBAUM, A. S. Sistemas Operacionais Modernos, LTC - Livros Técnicos e Científicos Editora S. A, Rio de Janeiro, 707p, 1995

TIPOS DE SOFTWARES DE COMPUTADORES. Disponível em: <http://www.ipv.cta.br/ead/anexo/12300.htm>, Acesso em: junho de 2004

TODAFRUTA. Cacau: Características gerais do cacau. Melhoramento Genético do Cacaueiro. Disponível em: <http://www.todafruta.com.br/todafruta/institucional.asp?menu=290>, Acesso em: 14/07/2004

UTZ, H. F. and A. E. MELCHINGER. PLABQTL: A program for composite interval mapping of QTL. *J. Quant. Trait Loci*, 2, 1996

VASCONCELOS, A. M. L de. Introdução à Engenharia de Software e aos Princípios de Qualidade. Curso de Pós-Graduação “Lato Sensu” a Distância – Melhoria de Processo de Software. Textos Acadêmicos. Editora UFLA (Universidade Federal de Lavras). Lavras, MG. 2004

WELLING, L. e THOMSON, L. PHP e MySQL Desenvolvimento WEB. Editora Campus, 676p, 2001

WOOLEY, J. C. e VARMA M. N. Computational biology: opportunity and challenges for the future. *Basic Life Sciences*, 63, 1-9, 1994

LEE, M. DNA markers in plant breeding programs. *Advances in Agronomy* 55, 265-344, 1995

MELCHINGER, A. E. Use of molecular markers in breeding of oligogenic disease resistance. *Plant Breeding*. 104, 1-19, 1990

YOUNG, N. D. A cautiously optimistic vision for marker-assisted breeding. *Molecular Breeding*. 5, 505-510, 1999

GLOSSÁRIO

Alelos: são as diferentes formas com que um gene pode se apresentar.

Algoritmo: é uma seqüência finita de passos lógicos e bem definidos para resolver um problema (descreve o padrão de comportamento). Esses passos posteriormente poderão ser escritos em uma linguagem de computador e transformados em um programa.

Banco de Dados Distribuídos: consiste em um banco de dados distribuído por uma coleção de computadores numa rede de computadores, i. e., a tecnologia de banco de dados fazendo uso dos recursos oferecidos pelas redes de computadores.

Bioinformática: uso de sistemas computacionais nas análises de caracterização molecular, integrando modelos matemáticos e estatísticos utilizados na interpretação e análise dos problemas biológicos, ou seja, é a ciência que usa informações para entender o funcionamento da biologia (GIBAS e JAMBECK, 2001).

Biotecnologia: é um conjunto de tecnologias que utilizam seres vivos para desenvolvimento de produtos e processos que tenham importância social e, ou, econômica como, por exemplo, na produção de pão, vinho, vacinas, dentre outros.

Cardinalidade: A cardinalidade é um número que expressa a freqüência (número de ocorrências) de determinada entidade associada a uma ocorrência da entidade em questão através do relacionamento; expressa o número de entidades às quais uma entidade pode ser associada via um conjunto de relacionamentos.

Chave Estrangeira: coluna ou combinação de colunas, cujos valores aparecem necessariamente na chave primária de uma tabela; mecanismo que permite a implementação de relacionamentos em um BD Relacional

Chave primária: é um campo chave para a estrutura de um BD, campo este que permite identificar os demais campos da estrutura.

Códigos mnemônicos: correspondem a abreviações semelhantes ao inglês, as quais representavam operações elementares do computador. chamadas de instruções de máquinas (TANEMBAUM, 1995). Como os códigos de máquina são binários ou hexadecimais, essas instruções compostas por três letras simplificam a tarefa de programação.

Compilador: programa utilizado para executar a tradução de uma linguagem para outra de mais baixo nível semântico. O objetivo de um compilador é partir de uma linguagem próxima à linguagem humana e gerar um programa equivalente em linguagem que a máquina entenda diretamente, criando o programa objeto, desde que o código fonte esteja coerente para execução.

Datawarehouse: (ou armazém de dados) é um sistema de computação utilizado para armazenar informações relativas às atividades de uma organização em bancos de dados, de forma consolidada. O desenho do BD favorece os relatórios e análise de grandes massas de dados e obtenção de informações estratégicas que podem facilitar a tomada de decisões.

DBA (Database Administrator): é o indivíduo que administra o banco de dados, de modo a determinar a melhor forma de organizar e armazenar dados em um banco de dados, enfim responsável por todas atividades de bancos de dados relativas por exemplo ao projeto, instalação e manutenção.

Escalabilidade: é a propriedade de um sistema qualquer que lhe confere a capacidade de aumentar seu desempenho sob carga quando recursos (tipicamente hardware, no caso de computadores) são acrescentados a esse sistema. A escalabilidade é um assunto extremamente importante em sistemas eletrônicos, bancos de dados, roteadores, redes de computadores, etc, e implica desempenho. Um sistema cujo desempenho aumenta com o acréscimo de hardware, proporcionalmente à capacidade acrescida, é chamado "sistema escalável".

Estrutura de dados: são agrupamentos de informações com um significado completo que normalmente são representados em estruturas de arranjos de dados, ou para as quais possam ser definidas operações próprias de conjuntos, tais como união e intersecção, ou ordenação e tratamento de repetição de valores (listas, filas, pilhas e montes).

Expressar: em genética, significa tornar possível de visualização; refere-se à característica fenotípica.

Fingerprinting: (impressão digital) é a descrição do padrão molecular de um indivíduo ou acesso.

Gene: é uma seqüência ou região do DNA cromossômico que possui uma função específica e consiste em uma seqüência específica de bases. Ele determina a síntese de uma proteína ou de um RNA estrutural ou de uma ribosima (RNA catalítico – Ribosima não é enzima; é um ácido nucléico com propriedades catalíticas).

Germoplasma: é a soma dos materiais genéticos de uma espécie. Assim, estes poderão ser na forma de plantas, anteras, pólen, sementes e tecidos (meristema, calo), células ou estruturas ainda mais simples; soma do material hereditário de uma espécie.

Heterozigoto: é o organismo que apresenta alelos diferentes em determinado loco gênico.

Interface: elemento pelo qual algum componente interage com outro, por exemplo, o usuário interage com um programa ou sistema operacional. Interação é: influência mútua de órgãos ou organismos inter-relacionados; ação recíproca de dois ou mais corpos; atividade ou trabalho compartilhado, em que existem trocas e influências recíprocas; comunicação entre pessoas que convivem; diálogo, trato, contato; intervenção e controle, feitos pelo usuário, do curso das atividades num programa de computador, num CD-ROM etc. (HOUAISS, 2004).

Integridade Referencial: integridade referencial, popularmente chamada de 'relacionamentos', é utilizada para garantir a integridade dos dados entre as tabelas relacionadas, promovendo obtenção de maior segurança nas informações. Ela assegura que um valor que aparece em uma relação (tabela) para um determinado conjunto de atributos apareça em outro conjunto de atributos em outra relação (tabela). Com a Integridade Referencial, o banco de dados não permite que seja cadastrado um pedido para um 'cliente' que ainda não foi cadastrado. Em outras palavras, ao cadastrar um pedido, o banco de dados verifica se o código do 'cliente' que foi digitado já existe na tabela 'Clientes'. Se não existir, o cadastro do pedido não será aceito.

Interpretador: programa utilizado que traduz código em uma linguagem de alto nível para a linguagem de máquina e roda uma linha por vez, até que todo programa seja executado. Por isso, pode ser visto como um tradutor que não gera código-objeto. A interpretação implica em mais tempo de execução, pois interpreta todo o comando toda a vez que vai executá-lo.

Iteração: é o ato de repetir um evento várias vezes. O processo iterativo consiste em encontrar a imagem de um ponto não apenas uma vez, mas repeti-lo interminavelmente usando sempre o último resultado obtido como ponto inicial para o próximo cálculo.

Loco gênico: é o local do cromossomo ocupado por um gene.

Marcador de DNA: segmento cromossômico geralmente separado por tamanho e que permite detectar diferenças entre dois ou mais indivíduos.

Marcador genético: uma proteína, ou um alelo, ou um fenótipo, usado para identificar um gene a ele ligado geneticamente.

Marcador molecular: proteína de reserva, ou isoenzima, ou fragmento de DNA que pode ser usado como marcador genético.

Marcas polimórficas: são regiões variáveis do DNA, cuja frequência na população é superior a 1% e inferior a 100%; marcas que variam entre os indivíduos em estudo. (Importante faz-se ressaltar que marcador não tem relação com gene, no sentido restrito dos termos).

Melhorista: pesquisador ou agrônomo especialista em Melhoramento Genético.

Middleware: é o neologismo criado para designar camadas de software que não constituem diretamente aplicações, mas que facilitam o uso de ambientes ricos em tecnologia da informação. A camada de middleware concentra serviços como identificação, autenticação, autorização, diretórios, certificados digitais e outras ferramentas para segurança.

Motif: seqüência do microssatélite que se repete. Ex. (AGC).

Multialélico: é o loco ou gene que apresenta mais de 2 tipos de alelos.

Mutações: alterações herdáveis resultante da alteração na seqüência de DNA.

Polimorfismo: é a existência de alelos diferentes na população num loco gênico.

Primer: é um segmento de DNA ou RNA de 15 a 40 bases que serve para iniciar o processo de amplificação do RNA ou DNA alvo; pequena seqüência de polinucleotídeo a qual desoxiribonucleotídeos podem ser adicionados pela DNA polimerase.

Propagação Vegetativa: multiplicação somática do indivíduo; reprodução que não envolve a união de gametas. A multiplicação pode se dar por borbulhas, colmos, estolões, rizomas, estacas etc.

Protocolo: em informática, significa o “apanhado de normas e especificações técnicas que regem a transmissão de dados entre computadores” (HOUAISS, 2004).

Recombinação: combinações de genes como resultado da segregação em cruzamentos de genitores geneticamente distintos. É também o rearranjo de genes ligados em virtude da permuta (*crossing over*).

Risco Ecológico: é o risco ambiental que pode haver de genes manipulados se espalharem pela natureza.

Suscetibilidade: inabilidade de uma planta para resistir, inibir ou evitar as atividades de um patógeno, praga, ou suportar uma condição adversa do ambiente.

Suscetível: organismo que não possui habilidade para resistir ao ataque de um patógeno ou inseto.

Tecnologia de Informação: é a aplicação de diferentes ramos da tecnologia no processamento de informações, sendo o meio pelo qual os dados são transformados e organizados para o uso das pessoas.

Tolerância: habilidade de uma planta em suportar o ataque de um patógeno ou praga sem expressiva redução da produtividade.

Variabilidade: estado de ser variável em qualquer categoria considerada.

Varietade crioula (Landrace): variedade não-melhorada, cultivada por produtores locais, originária de populações silvestres.

Varietade: categoria taxonômica de planta sempre abaixo daquela de espécie: 1) em taxonomia vegetal, a variedade ocupa uma posição abaixo da categoria de

subespécie, mas acima da categoria forma, e é sempre escrita em latim (exemplo: *Euphorbia milii* var. *milii*); 2) em melhoramento genético, a variedade é sinônimo de variedade cultivada e de cultivar.

Visões: trata-se da observação de segmentos do banco de dados criados a partir de consultas ao mesmo.

Vulnerabilidade Genética: condição de estreita diversidade genética, com elevado risco ecológico*.

APÊNDICES

APÊNDICE 1 - Casos de Uso

Manipular Dados de Acesso	
Ator: Usuário Pesquisador (iniciador)	
Finalidade: Atualizar os dados dos Acessos de Cacaueiro no BD.	
Visão Geral: Alterar dados do BD sobre Acessos, de acordo com suas permissões sobre esses dados.	
Seqüência Típica de Eventos	
Ação do Ator	Resposta do Sistema
1 – Este caso de uso começa quando um usuário acessa o formulário de interface para Acesso.	2 – Ver Seção: Ativar Formulário.
3 – Preencher o formulário e escolher uma opção disponível.	4 – Checar os dados do formulário e executar opção.
	5 – Exibir mensagem de falha ou sucesso de execução ao usuário.

Manipular Dados de Loco	
Ator: Usuário Pesquisador (iniciador)	
Finalidade: Atualizar os dados dos Locos no BD.	
Visão Geral: Alterar dados do BD sobre Locos, de acordo com suas permissões sobre esses dados.	
Seqüência Típica de Eventos	
Ação do Ator	Resposta do Sistema
1 – Este caso de uso começa quando um usuário acessa o formulário de interface para Loco.	2 – Ver Seção: Ativar Formulário.
3 – Preencher o formulário e escolher uma opção disponível.	4 – Checar os dados do formulário e executar opção.
	5 – Exibir mensagem de falha ou sucesso de execução ao usuário.

Ativar Formulário

Ator: Manipular Dados de Acesso (iniciador), Manipular Dados de Loco (iniciador) e Manipular Dados de <i>Primer</i> (iniciador)	
Finalidade: Disponibilizar para o usuário os serviços de Inserção, Deleção, Atualização e Impressão de informações do BD comuns aos formulários de interface.	
Visão Geral: Verificar permissões dos usuários e disponibilizar os serviços correspondentes.	
Seqüência Típica de Eventos	
Ação do Ator	Resposta do Sistema
1 – Este caso de uso começa quando um usuário preenche um campo especial do formulário de interface com o identificador de um registro do BD.	2 – Verificar através do identificador, a existência do registro no BD.
	3 – Recuperar as informações do BD, carregá-las no formulário de interface e disponibilizar as opções Impressão e Limpar Formulário.
	4 – Verificar se o Usuário autenticado é dono do registro localizado.
	5 – Disponibilizar opções de Deleção e Atualização (Salvar possível alteração feita).
Seqüência alternativa de eventos	
	2 – Disponibilizar opções Inserção e Limpar Formulário.

Manipular Dados de <i>Primer</i>	
Ator: Usuário Pesquisador (iniciador)	
Finalidade: Atualizar os dados dos <i>Primers</i> no BD.	
Visão Geral: Alterar dados do BD sobre <i>Primers</i> , de acordo com suas permissões sobre esses dados.	
Seqüência Típica de Eventos	
1 – Este caso de uso começa quando	2 – Ver Seção: Ativar Formulário.

um usuário acessa o formulário de interface <i>Primer</i> .	
3 – Preencher o formulário e escolher uma opção disponível.	4 – Checar os dados do formulário e executar opção.
	5 – Exibir mensagem de falha ou sucesso de execução ao usuário.

Cadastrar Artigo	
Ator: Usuário (iniciador)	
Finalidade: Cadastrar Artigo no BD.	
Visão Geral: Cadastrar Artigos Científicos no BD.	
Seqüência Típica de Eventos	
Ação do Ator	Resposta do Sistema
1 – Este caso de uso começa quando um usuário deseja cadastrar um Artigo Científico.	2 – Disponibilizar Formulário de interface para Artigo Científico.
3 - Preencher o formulário e escolher uma opção disponível. Se Autor não possuir cadastro, ver Seção: Cadastrar Autor	4 – Checar os dados do formulário e executar opção.
	5 – Exibir mensagem de falha ou sucesso de execução ao usuário.

Cadastrar Autor	
Ator: Cadastrar Artigo(iniciador), usuário	
Finalidade: Cadastrar Autores de Artigos Científicos no BD.	
Visão Geral: Usar o formulário de interface de Autor para cadastrar autores de artigos no BD.	
Seqüência Típica de Eventos	
Ação do Ator	Resposta do Sistema
1 – Este caso de uso começa quando um usuário necessita cadastrar um Autor.	2 – Disponibilizar Formulário de interface para Artigo.

3 - Preencher o formulário e escolher uma opção disponível.	4 – Checar os dados do formulário e executar opção.
	5 – Exibir mensagem de falha ou sucesso de execução ao usuário.

Cadastrar Usuário	
Ator: Gerente de Projeto (iniciador)	
Finalidade: Cadastrar um usuário do BD.	
Visão Geral: Cadastrar usuários do BD de modo a permitir inserção de dados no mesmo.	
Seqüência Típica de Eventos	
Ação do Ator	Resposta do Sistema
1 – Este caso de uso começa quando um gerente deseja conceder privilégio de usuário a um pesquisador.	2 – Disponibilizar Formulário de interface para Usuário.
3 – Preencher o formulário e escolher uma opção disponível.	4 – Checar os dados do formulário e executar opção.
	5 – Exibir mensagem de falha ou sucesso de execução ao gerente.

Autenticar usuário	
Ator: Visitante (iniciador)	
Finalidade: Permitir o acesso ao Sistema de Informação aos usuários cadastrados.	
Visão Geral: Verificar do usuário a permissão de edição de dados no BD através de identificação única e pessoal.	
Seqüência Típica de Eventos	
Ação do Ator	Resposta do Sistema
1 – Este caso de uso começa quando um visitante deseja se identificar com o sistema.	2 – Solicita <i>login</i> e senha ao visitante para verificar suas permissões.
	3 – Permitir acesso aos formulários de edição do BD.

Seqüência alternativa de eventos	
	2 – Negar permissão ao visitante.

Cadastrar Instituição	
Ator: Gerente do Projeto (iniciador)	
Finalidade: Cadastrar Instituição no BD.	
Visão Geral: Cadastrar Instituições no BD.	
Seqüência Típica de Eventos	
Ação do Ator	Resposta do Sistema
1 – Este caso de uso começa quando o gerente de projeto necessita cadastrar uma Instituição.	2 – Disponibilizar Formulário de interface para Instituição.
3 - Preencher o formulário e escolher uma opção disponível.	4 – Checar os dados do formulário e executar opção.
	5 – Exibir mensagem de falha ou sucesso de execução ao usuário.

Cadastrar Laboratório	
Ator: Gerente do Projeto (iniciador)	
Finalidade: Cadastrar um Laboratório de Pesquisa no BD.	
Visão Geral: Cadastrar um Laboratório de Pesquisa que aloca usuários do BD.	
Seqüência Típica de Eventos	
Ação do Ator	Resposta do Sistema
1 – Este caso de uso começa quando o gerente de projeto necessita cadastrar um Laboratório.	2 – Disponibilizar Formulário de interface para Laboratório.
3 – Preencher o formulário e escolher uma opção disponível.	4 – Checar os dados do formulário e executar opção.
	5 – Exibir mensagem de falha ou sucesso de execução ao usuário.

Criar grupo de trabalho

Ator: Usuário (iniciador)	
Finalidade: Cadastrar um grupo de trabalho no BD.	
Visão Geral: Cadastrar um grupo de trabalho no BD formado por usuários e que permita o compartilhamento de dados no BD.	
Seqüência Típica de Eventos	
Ação do Ator	Resposta do Sistema
1 – Este caso de uso começa quando o usuário deseja criar um grupo de trabalho para compartilhar os dados produzidos com seus integrantes.	2 – Disponibilizar Formulário de interface para Grupo de Trabalho.
3 – Preencher o formulário e escolher uma opção disponível.	4 – Checar os dados do formulário e executar opção.
	5 – Exibir mensagem de falha ou sucesso de execução ao usuário.

Realizar Consultas Avançadas	
Ator: Visitante (iniciador)	
Finalidade: Criar relatórios a partir de consultas personalizadas.	
Visão Geral: Criar relatórios a partir da seleção de opções disponíveis no formulário de interface de Relatórios.	
Seqüência Típica de Eventos	
Ação do Ator	Resposta do Sistema
1 – Este caso de uso começa quando um visitante deseja um relatório personalizado.	2 – Disponibilizar Formulário de interface para Criação de Consultas.
3 – O visitante preenche o formulário e escolhe uma opção disponível.	4 – Checar os dados do formulário e gerar fielmente a consulta ao BD.
	5 – Montar relatório e exibi-lo ao visitante com as opções Impressão e <i>Download</i> .

Pesquisar Artigos

Ator: Visitante (iniciador)	
Finalidade: Pesquisar o material bibliográfico existente no BD.	
Visão Geral: Pesquisar artigos científicos cadastrados no BD.	
Seqüência Típica de Eventos	
Ação do Ator	Resposta do Sistema
1 – Este caso de uso começa quando um visitante deseja pesquisar artigos científicos no BD.	2 – Disponibilizar formulário de interface com as opções de pesquisa de artigos científicos.
3 – O visitante preenche o formulário e escolhe uma opção disponível.	4 – Checar os dados do formulário e gerar fielmente a consulta ao BD.
	5 – Exibir resultados da pesquisa com opção para <i>download</i> .

APÊNDICE 2 – Projeto Físico do banco de dados.

1 – A TABELA ACESSO: Armazena os dados sobre os acessos de cacauero analisados em laboratório. Possui os seguintes campos:

idAcesso: O próprio nome que o identifica.

origem: Referência à localização geográfica de onde o acesso foi obtido.

corFruto: Cor do fruto.

altura: Altura média.

resistência: Informa o nível de resistência à vassoura-de-bruxa, podendo ser: alto, médio ou baixo.

frutosPorArroba: média de frutos necessários para se alcançar uma arroba.

pesoFrutoSeco: Peso médio de um fruto seco.

gorduraSemente: Peso médio de gordura de uma semente.

pesoPolpaFruto: Peso médio da polpa do fruto.

pesoCemSementes: Peso médio de 100 sementes secas.

compatibilidade: Informa se a planta é auto-compatível ou incompatível.

casca: Informa se casca é lisa ou rugosa.

primeiraColheita: tempo médio esperado para realizar a primeira colheita do cultivar.

dataInsercao: Data de inserção dos Dados.

dataAtualizacao: Data em que ocorreu a última atualização dos dados.

ultimoAeditar: Usuário do grupo que foi o último a modificar algum dado.

2 – A TABELA LOCO: Armazena os dados sobre os locos genéticos encontrados em laboratório. Possui os seguintes campos:

idLoco: O próprio nome que o identifica.

acesso: Em qual acesso foi encontrado o alelo.

tamAlelo1: Tamanho do alelo 1.

tamAlelo2: Tamanho do alelo 2.

dataInsercao: Data de inserção dos Dados.

dataAtualizacao: Data em que ocorreu a última atualização dos dados.

ultimoAeditar: Usuário do grupo que foi o último a modificar algum dado.

3 – A TABELA PRIMER: Armazena os dados sobre os *primers* empregados em laboratório. Possui os seguintes campos:

idPrimer: Nome que o identifica.

marcador: Tipo de marcador do *primer*.

sequencia: Sua própria seqüência.

tempAnelamento: Temperatura de anelamento do *primer*.

grupoDeLigacao: Grupo de ligação que o representa.

motify: Informa o número de repetições das bases do *primer*.

dono: Informa qual grupo é dono das informações sobre o Loco.

dataInsercao: Data de inserção dos Dados.

dataAtualizacao: Data em que ocorreu a última atualização dos dados.

ultimoAeditar: Usuário do grupo que foi o último a modificar algum dado.

4 – A TABELA AUTOR: Armazena os dados sobre os autores dos artigos científicos. Possui os seguintes campos:

idAutor: Identificador do registro autor.

nome: Nome completo do autor.

e-mail: endereço eletrônico da instituição.

5 – A TABELA LISTADEPALAVRAS: Armazena as palavras que são palavras-chave referenciadas num registro da tabela Artigo. Possui o seguinte campo:

palavra: palavra usada.

6 – A TABELA TIPOMARCADOR: Armazena os tipos de marcadores usados em laboratório. Possuem os seguintes campos:

idTipoMarcador: Identificador do registro TipoMarcador.

tipo: Próprio tipo de marcador

7 – A TABELA ARTIGO: Armazena dados sobre os artigos científicos produzidos e que referenciam um registro da tabela Loco. Possuem os seguintes campos:

idArtigo: Identificador do registro artigo.

titulo: Título do artigo científico.

revistaPublicacao: Nome da revista em que foi publicado o artigo.

anoPublicacao: Ano de publicação do artigo.

resumo: Resumo da publicação.

dataInsercao: Data de inserção dos Dados.

dataAtualizacao: Data em que ocorreu a última atualização dos dados.

ultimoAeditar: Usuário do grupo que foi o último a modificar algum dado.

8 – A TABELA ARTIGO_HAS_LOCO: Armazena dados da relação entre as tabelas Artigo e Loco. Possuem os seguintes campos:

loco: Referência ao Loco publicado.

artigo: Referência ao Artigo que publica o loco.

9 – A TABELA AUTOR_HAS_ARTIGO: Armazena dados da relação entre as tabelas Autor e Artigo. Possuem os seguintes campos:

autor: Referência ao Autor de um artigo.

artigo: Referência ao Artigo publicado.

10 – A TABELA LISTADEPALAVRAS_HAS_ARTIGO: Armazena dados da relação entre as tabelas ListaDePalavras e Artigo. Possuem os seguintes campos:

palavra: Referência às palavras que são palavras-chave de um artigo.

artigo: Referência ao Artigo científico.

11 – A TABELA INSTITUICAO: Armazena dados de Instituições de Pesquisa envolvidas no programa de melhoramento genético. Possuem os seguintes campos:

idInstituicao: Identificador do registro Instituição.

nome: Nome da Instituição.

endereco: Endereço da Instituição.

12 – A TABELA LABORATORIO: Armazena dados dos laboratórios envolvidos no programa de melhoramento genético. Possuem os seguintes campos:

idLaboratorio: Identificador do registro Laboratório.

instituicao: Referência a Instituição que o laboratório faz parte.

nome: Nome do laboratório.

endereco: Endereço do laboratório.

13 – A TABELA PESQUISADOR: Armazena dados sobre os usuários com permissão de manipular dados do banco de dados. Possuem os seguintes campos:

usuario: Nome de usuário reconhecido pelo sistema.

senha: Senha de identificação.

nome: Nome completo do pesquisador.

laboratorio: Laboratório em que é alocado.

email: E-mail para contato.

fone: Telefone de contato.

14 – A TABELA GRUPODEPESQUISA: Armazena dados sobre os grupo formado por usuário e com permissão de compartilhar seus respectivos dados inseridos no banco de dados. Possuem os seguintes campos:

nome: Nome que o identifica.

pesquisadorResponsável: pesquisador responsável.

15 – A TABELA PESQUISADOR_HAS_GRUPODEPESQUISA: Armazena dados da relação entre as tabelas Grupo e Usuario. Possuem os seguintes campos:

grupo: Referência ao grupo criado.

usuario: Referência ao usuário integrante.

16 – A TABELA PESQUISA: Armazena dados da identificação de Pesquisas Científicas. Possuem os seguintes campos:

nome: Nome que a identifica.

grupo: Referência ao grupo que a mantém.

dataInsercao: Data de inserção dos Dados.

17 – A TABELA PESQUISA_HAS_LOCO: Armazena dados da relação entre as tabelas Pesquisa e Loco. Possuem os seguintes campos:

pesquisa: Referência a pesquisa criada.

loco: Referência ao Loco marcado.

APÊNDICE 3 - Código gerador do Banco de Dados em MySQL.

```
CREATE TABLE Instituicao (  
  idInstituicao INTEGER UNSIGNED NOT NULL AUTO_INCREMENT,  
  nome VARCHAR(100) NULL,  
  endereco VARCHAR(255) NULL,  
  PRIMARY KEY(idInstituicao)  
)  
TYPE=InnoDB;
```

```
CREATE TABLE ListaDePalavra (  
  palavra VARCHAR(45) NOT NULL,  
  PRIMARY KEY(palavra)  
)  
TYPE=InnoDB;
```

```
CREATE TABLE TipoMarcador (  
  idTipoMarcador INTEGER UNSIGNED NOT NULL AUTO_INCREMENT,  
  tipo VARCHAR(15) NULL,  
  PRIMARY KEY(idTipoMarcador)  
)  
TYPE=InnoDB;
```

```
CREATE TABLE Autor (  
  idAutor INTEGER UNSIGNED NOT NULL AUTO_INCREMENT,  
  nome VARCHAR(255) NULL,  
  PRIMARY KEY(idAutor)  
)  
TYPE=InnoDB;
```

```
CREATE TABLE Artigo (  
  idArtigo INTEGER UNSIGNED NOT NULL AUTO_INCREMENT,  
  titulo VARCHAR(255) NULL,  
  revistaPublicacao VARCHAR(255) NULL,  
  anoPublicacao CHAR(4) NULL,  
  resumo BLOB NULL,  
  dataInsercao DATE NULL,  
  dataAtualizacao DATE NULL,  
  ultimoAeditar VARCHAR(45) NULL,  
  PRIMARY KEY(idArtigo)  
)  
TYPE=InnoDB;
```

```
CREATE TABLE Acesso (  
  idAcesso VARCHAR(45) NOT NULL,  
  origem VARCHAR(200) NULL,  
  corFruto VARCHAR(25) NULL,  
  altura FLOAT NULL,
```

```

resistencia VARCHAR(5) NULL,
frutosPorArroba FLOAT NULL,
pesoFrutoSeco FLOAT NULL,
gorduraSemente FLOAT NULL,
pesoPolpaFruto FLOAT NULL,
pesoCemSementes FLOAT NULL,
compatibilidade VARCHAR(13) NULL,
casca VARCHAR(6) NULL,
primeiraColheita VARCHAR(2) NULL,
dataAtualizacao DATE NULL,
ultimoAeditar VARCHAR(45) NULL,
dataInsercao DATE NULL,
PRIMARY KEY(idAcesso)
)
TYPE=InnoDB;

CREATE TABLE Primer (
idPrimer VARCHAR(45) NOT NULL,
marcador INTEGER UNSIGNED NOT NULL,
sequencia VARCHAR(45) NULL,
tempAnelamento FLOAT NULL,
motify VARCHAR(10) NULL,
grupoDeLigacao VARCHAR(45) NULL,
dataInsercao DATE NULL,
dataAtualizacao DATE NULL,
ultimoAeditar VARCHAR(45) NULL,
PRIMARY KEY(idPrimer),
INDEX TipoMarcador_Primer(marcador),
FOREIGN KEY(marcador)
REFERENCES TipoMarcador(idTipoMarcador)
ON DELETE NO ACTION
ON UPDATE NO ACTION
)
TYPE=InnoDB;

CREATE TABLE Laboratorio (
idLaboratorio INTEGER UNSIGNED NOT NULL AUTO_INCREMENT,
instituicao INTEGER UNSIGNED NOT NULL,
nome VARCHAR(255) NULL,
endereco VARCHAR(255) NULL,
PRIMARY KEY(idLaboratorio),
INDEX Laboratorio_Instituicao(instituicao),
FOREIGN KEY(instituicao)
REFERENCES Instituicao(idInstituicao)
ON DELETE NO ACTION
ON UPDATE NO ACTION
)
TYPE=InnoDB;

```

```

CREATE TABLE Autor_has_Artigo (
  autor INTEGER UNSIGNED NOT NULL,
  artigo INTEGER UNSIGNED NOT NULL,
  PRIMARY KEY(autor, artigo),
  INDEX Autor_has_Artigo_FKIndex1(autor),
  INDEX Autor_has_Artigo_FKIndex2(artigo),
  FOREIGN KEY(autor)
    REFERENCES Autor(idAutor)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
  FOREIGN KEY(artigo)
    REFERENCES Artigo(idArtigo)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION
)
TYPE=InnoDB;

```

```

CREATE TABLE Loco (
  idLoco VARCHAR(45) NOT NULL,
  acesso VARCHAR(45) NOT NULL,
  tamAlelo1 INTEGER UNSIGNED NULL,
  tamAlelo2 INTEGER UNSIGNED NULL,
  dataInsercao DATE NULL,
  dataAtualizacao DATE NULL,
  ultimoAeditar VARCHAR(45) NULL,
  PRIMARY KEY(idLoco),
  INDEX Acesso_Loco(acesso),
  INDEX Primer_Loco(idLoco),
  FOREIGN KEY(acesso)
    REFERENCES Acesso(idAcesso)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
  FOREIGN KEY(idLoco)
    REFERENCES Primer(idPrimer)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION
)
TYPE=InnoDB;

```

```

CREATE TABLE ListaDePalavra_has_Artigo (
  artigo INTEGER UNSIGNED NOT NULL,
  palavra VARCHAR(45) NOT NULL,
  PRIMARY KEY(artigo, palavra),
  INDEX ListaDePalavra_has_Artigo_FKIndex1(palavra),
  INDEX ListaDePalavra_has_Artigo_FKIndex2(artigo),
  FOREIGN KEY(palavra)
    REFERENCES ListaDePalavra(palavra)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,

```

```
FOREIGN KEY(artigo)
REFERENCES Artigo(idArtigo)
ON DELETE NO ACTION
ON UPDATE NO ACTION
```

```
)
TYPE=InnoDB;
```

```
CREATE TABLE Pesquisador (
usuario VARCHAR(45) NOT NULL,
laboratorio INTEGER UNSIGNED NOT NULL,
senha VARCHAR(20) NULL,
nome VARCHAR(255) NULL,
mail VARCHAR(45) NULL,
fone VARCHAR(15) NULL,
PRIMARY KEY(usuario),
INDEX Pesquisador_FKIndex1(laboratorio),
FOREIGN KEY(laboratorio)
REFERENCES Laboratorio(idLaboratorio)
ON DELETE NO ACTION
ON UPDATE NO ACTION
```

```
)
TYPE=InnoDB;
```

```
CREATE TABLE GrupoDePesquisa (
nome VARCHAR(45) NOT NULL,
pesquisadorResponsavel VARCHAR(45) NOT NULL,
PRIMARY KEY(nome),
INDEX GrupoDePesquisa_FKIndex1(pesquisadorResponsavel),
FOREIGN KEY(pesquisadorResponsavel)
REFERENCES Pesquisador(usuario)
ON DELETE NO ACTION
ON UPDATE NO ACTION
```

```
)
TYPE=InnoDB;
```

```
CREATE TABLE Pesquisador_has_GrupoDePesquisa (
grupo VARCHAR(45) NOT NULL,
pesquisador VARCHAR(45) NOT NULL,
PRIMARY KEY(grupo, pesquisador),
INDEX Pesquisador_has_GrupoDePesquisa_FKIndex1(pesquisador),
INDEX Pesquisador_has_GrupoDePesquisa_FKIndex2(grupo),
FOREIGN KEY(pesquisador)
REFERENCES Pesquisador(usuario)
ON DELETE NO ACTION
ON UPDATE NO ACTION,
FOREIGN KEY(grupo)
REFERENCES GrupoDePesquisa(nome)
ON DELETE NO ACTION
ON UPDATE NO ACTION
```

);

```
CREATE TABLE Artigo_has_Loco (  
  loco VARCHAR(45) NOT NULL,  
  artigo INTEGER UNSIGNED NOT NULL,  
  PRIMARY KEY(loco, artigo),  
  INDEX Artigo_has_Loco_FKIndex1(artigo),  
  INDEX Artigo_has_Loco_FKIndex2(loco),  
  FOREIGN KEY(artigo)  
    REFERENCES Artigo(idArtigo)  
    ON DELETE NO ACTION  
    ON UPDATE NO ACTION,  
  FOREIGN KEY(loco)  
    REFERENCES Loco(idLoco)  
    ON DELETE NO ACTION  
    ON UPDATE NO ACTION  
)  
TYPE=InnoDB;
```

```
CREATE TABLE Pesquisa (  
  nome VARCHAR(100) NOT NULL,  
  grupo VARCHAR(45) NOT NULL,  
  dataInsercao DATE NULL,  
  PRIMARY KEY(nome),  
  INDEX Pesquisa_FKIndex1(grupo),  
  FOREIGN KEY(grupo)  
    REFERENCES GrupoDePesquisa(nome)  
    ON DELETE NO ACTION  
    ON UPDATE NO ACTION  
)  
TYPE=InnoDB;
```

```
CREATE TABLE Pesquisa_has_Loco (  
  pesquisa VARCHAR(100) NOT NULL,  
  loco VARCHAR(45) NOT NULL,  
  PRIMARY KEY(pesquisa, loco),  
  INDEX Pesquisa_has_Loco_FKIndex1(pesquisa),  
  INDEX Pesquisa_has_Loco_FKIndex2(loco),  
  FOREIGN KEY(pesquisa)  
    REFERENCES Pesquisa(nome)  
    ON DELETE NO ACTION  
    ON UPDATE NO ACTION,  
  FOREIGN KEY(loco)  
    REFERENCES Loco(idLoco)  
    ON DELETE NO ACTION  
    ON UPDATE NO ACTION  
);
```

APÊNDICE 4 – Formulário Semi-estruturado



**UNIVERSIDADE ESTADUAL DE SANTA CRUZ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE CIÊNCIAS BIOLÓGICAS
MESTRADO EM GENÉTICA E BIOLOGIA MOLECULAR**

Pesq.

FORMULÁRIO SEMI-ESTRUTURADO PARA ENTREVISTA

A realização do projeto “Estabelecimento de Sistema Automático e Informatizado para Análises de Diversidade Genética e Genotipagem de Acessos de Cacaueiro *Theobroma cacao L.*” será de grande proveito para as instituições de pesquisa e agricultores, colaborando para melhor utilização das informações e conhecimentos gerados.

Este formulário visa obter informações que possam auxiliar na definição da estrutura e das características de um banco de dados de marcadores genéticos do cacau e os procedimentos para sua manipulação. Essas informações também serão úteis na definição de estrutura e funções de uma interface WEB para acesso ao banco.

1. Do resultado final do banco de dados (BD):

O que você deseja utilizar do BD?

2. Informações referentes ao BD

2.1. Quais as informações agronômicas da planta devem estar armazenadas no BD?

Dados Armazenados	Tipo do dado
cor do fruto	texto (quantos caracteres)
altura da planta	nº decimal (quantas casas decimais)
nível de resistência a VB	alto/médio/baixo → nº p/ cálculo ou texto?
nº de frutos por arroba	nº (quantas casas)

quantidade de gordura por semente	nº decimal ou porcentagem?
quantidade de polpa por fruto	nº decimal ou porcentagem?
peso fruto seco	nº decimal (quantas casas decimais)
peso semente seca (cada 100)	nº decimal (quantas casas decimais)
tempo para primeira colheita	nº decimal (quantas casas decimais)
local de origem	texto / quantos caracteres?
nome / código	texto / quantos caracteres?
outros atributos relevantes:	

2.2. Quais dados sobre descritores morfológicos e fisiológicos das plantas devem estar armazenados no BD?

2.3. Lista de acessos em anexo. Para contemplar as necessidades da comparação é necessário existir um grupo de dados que sejam representativos para tal fim. Destes materiais da lista quais você identifica como sendo pertinentes para formar um grupo de referência para análises comparativas ?

2.4. Quais dados moleculares você julga necessário existir no BD para efetuar análise genética?

2.5. Quais dados moleculares você julga necessário existir no BD para efetuar análise da relação de parentesco?

2.6. Quais dados moleculares você julga necessário existir no BD para efetuar análise de fluxo gênico?

3.3. Gostaria de:

- a. Consultar sites relacionados com genoma não sim
- b. Incluir publicações sobre cacau não sim
- c. Incluir marcadores presentes em acessos de cacau não sim
- d. Outras necessidades para este site:

3.4. Neste site serão disponibilizados os programas para análises genéticas de marcadores moleculares.

- a. Prefere utilizar on-line ou fazer o download dos programas?
 on-line download
- b. Você utiliza o programa GENES? sim não
- c. Você utiliza o programa GQMOL? sim não
- d. Quais outros programas de análise você utiliza?

4. Outras Informações / Observações:



**UNIVERSIDADE ESTADUAL DE SANTA CRUZ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE CIÊNCIAS BIOLÓGICAS
MESTRADO EM GENÉTICA E BIOLOGIA MOLECULAR**

Melhor.

FORMULÁRIO SEMI-ESTRUTURADO PARA ENTREVISTA

A realização do projeto “Estabelecimento de Sistema Automático e Informatizado para Análises de Diversidade Genética e Genotipagem de Acessos de Cacaueiro *Theobroma cacao L.*” será de grande proveito para as instituições de pesquisa e agricultores, colaborando para melhor utilização das informações e conhecimentos gerados.

Este formulário visa obter informações que possam auxiliar na definição da estrutura e das características de um banco de dados de marcadores genéticos do cacau e os procedimentos para sua manipulação. Essas informações também serão úteis na definição de estrutura e funções de uma interface WEB para acesso ao banco.

1. Do resultado final do banco de dados (BD):

O que você deseja utilizar do BD?

2. Informações referentes ao BD

2.1. Quais as informações agronômicas da planta devem estar armazenadas no BD?

Dados Armazenados	Tipo do dado
cor do fruto	texto (x caracteres)
altura da planta	nº decimal (x casas decimais)
nível de resistência a VB	alto/médio/baixo→ nº p/ cálculo ou texto?
nº de frutos por arroba	nº (x casas)
quantidade de gordura por semente	nº decimal ou porcentagem?
quantidade de polpa por fruto	nº decimal ou porcentagem?
peso fruto seco	nº decimal (x casas decimais)
peso semente seca (cada 100)	nº decimal (x casas decimais)
tempo para primeira colheita	nº decimal (x casas decimais)
local de origem	texto / quantos caracteres?
nome / código	texto / quantos caracteres?
outras características relevantes:	

2.2. Quais dados sobre descritores morfológicos e fisiológicos das plantas devem estar armazenados no BD?

2.3. Lista de acessos em anexo. Para contemplar as necessidades da comparação é necessário existir um grupo de dados que sejam representativos para tal fim. Destes materiais da lista quais você identifica como sendo pertinentes para formar um grupo de referência para análises comparativas ?

2.4. Quais dados moleculares você deseja que exista no BD para efetuar análise genética?

2.5. Quais dados moleculares você deseja que exista no BD para efetuar análise da relação de parentesco?

2.6. Quais dados moleculares você deseja que exista no BD para efetuar análise de fluxo gênico?

2.7. Quais os produtos da análise e respectivas características relevantes devem estar presentes no BD ?

Dados Gerados	Tipo do dado
coeficiente de dissimilaridade	nº decimal (x casas decimais)
localização do cromossomo	nº decimal (x casas decimais)
freqüência alélica	nº decimal (x casas decimais)

2.8. Quais as características importantes de um *primer*?

Atributos	Tipo do dado
seqüência do <i>primer</i>	texto / pares de base (de 10 a 35 pb)
temperatura de anelamento	nº decimal (x casas decimais) / até 72°C

2.9. Quais os atributos de um alelo?

Atributos	Tipo do dado
seqüência do alelo	texto / pares de base (de 10 a 35 pb)
nome	texto / quantos caracteres?
freqüência alélica	nº decimal (x casas decimais)
tamanho	texto / pares de base (qual limite?)

3.4. Neste site serão disponibilizados os programas para análises genéticas de marcadores moleculares.

a. Prefere utilizar on-line ou fazer o download dos programas?

b. on-line download

c. Você utiliza o programa GENES? sim não

d. Você utiliza o programa GQMOL? sim não

e. Quais outros programas de análise você utiliza?

4. Outras Informações / Observações:

APÊNDICE 5 – Lista de acessos de cacau

LISTA DE ACESSOS DE CACAU

1	SCA6	18	SJ02	35	AM-1
2	SIAL169	19	SP50	36	AM-2
3	CEPEC42	20	ALMC5	37	AM-3
4	COCA	21	PA300	38	M-4
5	NO02	22	SR03	39	M-5
6	VB1151	23	EET392	40	T-11
7	SIAL505	24	RB36	41	T-10
8	VB1154	25	SIC328	42	FL60
9	VB1117	26	SIC329	43	FL77
10	LCTEEN162-1010	27	SIAL325	44	TSH516
11	FB21	28	UF168	45	SCA12
12	GNV(II)	29	IMC67	46	SIAL70
13	AMAZON15	30	PA30	47	SCA12
14	VB1139	31	RB39	48	EEG29
15	LCTEEN37A	32	TSA644	49	EET397
16	TSH1188	33	Cruz.do Sul 03	50	TSH565
17	SJ01	34	CCN51		