Organização e estruturação em XML das características morfológicas de cromossomos a partir de uma base de imagens para o processo de identificação de cromossomos

Guilherme Chagas Kurtz¹
Alessandro Mainardi de Oliveira¹
Alexandre de Oliveira Zamberlan¹
Reiner Franthesco Perozzo¹
Ricardo Frohlich da Silva¹

Resumo: o objetivo deste trabalho é apresentar uma simplificação, estruturação e organização das informações a respeito das características morfológicas de cromossomos a partir de uma base de imagens, aplicados ao processo de identificação automática de cromossomos humanos. Neste trabalho, foram realizadas pesquisas com o intuito de investigar como funciona o processo de identificação de cromossomos humanos, quais são as principais informações morfológicas utilizadas bem como as opções já existentes e utilizadas por outros autores, analisando suas características e fragilidades. Por fim, são apresentadas duas propostas de estruturação e organização dessas informações utilizando XML Schema e XML.

Palavras-chave: Cromossomos. XML. Características morfológicas. Estruturação. Identificação de cromossomos.

Abstract: The objective of this paper is to present a simplification, structuring and organization of informations about the morphological characteristics of chromosomes from an image dataset, applied to the automatic process of identifying human chromosomes. In this paper, surveys were conducted in order to investigate how the process of identifying human chromosomes works and what are the main morphological information used in it, as well as existing options used by other authors, analyzing their characteristics and weaknesses. Finally, we present two proposals for structuring and organizing such information using XML Schema and XML.

Keywords: Chromosomes. XML. Morphological characteristics. Structuring. Chromosomes identification.

1 Introdução

Nas últimas duas décadas, a área da genética, em especial a genética médica, cresceu muito no Brasil e tem atraído um grande número de profissionais. A citogenética humana foi uma de suas primeiras subáreas a ser implantada no país, inicialmente em laboratórios de pesquisa e, mais recentemente, também em laboratórios de análises clínicas. Atualmente, suas aplicações incluem a caracterização de polimorfismos nas populações e a pesquisa da ação de agentes mutagênicos (que causam mutações ou mudanças na forma do DNA) ou carcinogênicos (que causam mutações e levam a ativação de genes tumorais) em ensaios *in vitro*, além da análise de cariótipo em muitas doenças. Suas técnicas compreendem importantes ferramentas de diagnóstico pré e pósnatal de anomalias congênitas (alterações nas quais os indivíduos já nascem com ela) e de diagnóstico e monitoramento de terapia em casos de neoplasias, principalmente hematológicas [1].

{guilhermekurtz,alexz,reiner.perozzo,Ricardo.frohlich@unifra.br}

http://dx.doi.org/10.5335/rbca.v8i2.5251

¹ Curso de Sistema de Informação, Centro Universitário Franciscano, Conjunto 1 − Rua dos Andradas 1614 − Santa Maria (RS) - Brasil

Os avanços ocorridos nas técnicas de cultura celular, bandeamento, coleta e análise dos materiais para a execução do cariótipo possibilitaram grandes progressos no diagnóstico das alterações cromossômicas, mas apesar da demanda crescente por esse tipo de exame, é pequena a oferta de sistemas automáticos que auxiliem o trabalho dos geneticistas na geração do cariótipo [1]. A possibilidade de se obter resultado em curto espaço de tempo, agilizando condutas terapêuticas ou tranquilizando a família, tem um valor inestimável. Assim, o cariótipo no diagnóstico clínico é uma verdadeira corrida contra o tempo [2].

Portanto, o desenvolvimento de um sistema computadorizado utilizando técnicas de visão por computador pode servir de auxílio ao geneticista tanto na otimização do trabalho quanto na execução da análise da forma dos cromossomos humanos. Por meio da automatização do processo de identificação dos cromossomos e elaboração do cariótipo, é possível aumentar a homogeneidade dos resultados da análise dos cromossomos, permitindo sua análise detalhada e a impressão das imagens digitalizadas dos cariótipos, que poderá ser enviada juntamente com os resultados. A meta dos laboratórios que prestam serviços nessa área deve ser o desenvolvimento de uma política de qualidade que faça com que o número de erros tenda a zero e que os laudos emitidos sejam compreensíveis pelos profissionais que atuam com o paciente.

Apesar disso, o desenvolvimento desse tipo de sistema é uma abordagem complexa, pois passa por diversas etapas e exige o conhecimento de diferentes áreas da computação, tais como processamento de imagens e inteligência artificial. Dentre essas etapas, a obtenção de uma base de informações a respeito das características morfológicas dos cromossomos é essencial, para que o sistema consiga identificar automaticamente os cromossomos presentes nas imagens obtidas por meio de um microscópio. Geralmente, essa base de informações é representada por uma base de imagens contendo centenas de amostras de cada tipo de cromossomo. As novas entradas são então comparadas com essas imagens de maneira que torne possível a identificação dos cromossomos dessa nova imagem. Esse processo pode se tornar muito complexo e pouco eficiente, pois se torna necessário, além da obtenção de uma base de imagens, o desenvolvimento de algoritmos de extração de informações dessas imagens e também de treinamento do sistema com base nessas informações. De maneira a simplificar esse processo, surge a proposta deste trabalho.

Dessa forma, o objetivo desse trabalho é realizar a organização e estruturação das características morfológicas de cromossomos humanos a partir de uma base de imagens, cujas características são utilizadas no processo de identificação de cromossomos, e então descrevê-las e modelá-las utilizando XML (eXtensible Markup Language) e XML Schema, que é uma linguagem baseada no formato XML utilizada para definição de regras de validação de documentos XML [3]. Por meio dessa simplificação, qualquer sistema de identificação de cromossomos e de elaboração do cariótipo pode fazer uso dessa base de informações sem a necessidade de se preocupar com a obtenção de imagens para treinamento do sistema bem como com o desenvolvimento de algoritmos de processamento de imagens para a extração dessas características ou de algoritmos de treinamento. Além disso, essas informações, por terem sido modeladas em XML, permitem que o acesso à elas ocorra de forma padronizada, assim como a alimentação com novas informações.

Este artigo está organizado da seguinte forma: a seção 2 apresenta uma breve revisão bibliográfica sobre o processo de elaboração do cariótipo. A seção 3 apresenta a metodologia utilizada para o desenvolvimento do trabalho e a seção 4 mostra os detalhes de como as informações a respeito das características morfológicas dos cromossomos foram representadas em XML, além dos resultados obtidos. Por fim, na seção 5, encontra-se a conclusão deste trabalho.

2 Fundamentação teórica

Nesta seção, serão feitos os levantamentos bibliográficos necessários para a compreensão do tema tratado neste trabalho.

2.1 Processo de elaboração do cariótipo

O processo de análise e identificação de cromossomos é chamado cariotipagem. A Figura 1.a traz um exemplo de uma metáfase de uma célula humana (conjunto de cromossomos em uma célula) e a Figura 1.b um exemplo de metáfase já identificada, em que o processo de identificação dos cromossomos já foi feito (cariotipagem). Esse processo é geralmente utilizado como recurso para investigação de alterações cromossômicas que podem ser responsáveis por diversos problemas de saúde.

Dessas alterações, podemos citar neoplasias (cânceres), síndromes genéticas, quebras cromossômicas (causadas por radiação, por exemplo). Além disso, é possível dar um prognóstico ao paciente, ou seja, definir a evolução de uma doença em certo organismo (nesse caso, em humanos) e a seguir descrever-lhe a conduta terapêutica necessária [4].

(1a) (1b)

Figura 1: Metáfase (a) e cariótipo (b) de células humanas

Fonte: elaborado pelos autores com dados da pesquisa.

Geralmente, esse processo de análise e identificação dos cromossomos é feito de forma manual, por um geneticista, que pode levar em torno de meia hora para estudar cada imagem. O processo de identificação manual envolve as seguintes etapas:

- a) aquisição da imagem Geralmente por intermédio de uma câmera acoplada a um microscópio;
- b) segmentação e identificação dos cromossomos Processo feito de forma manual, em que se desenham as metáfases (no mínimo cinquenta metáfases por paciente), identificando e analisando cada cromossomo, e em seguida as metáfases são impressas e recortadas manualmente (segmentação);
- c) identificação de anomalias Com base em cromossomos já identificados, busca-se por anomalias que podem ocorrer. Tais anomalias podem estar relacionadas tanto ao número de cromossomos quanto a alterações estruturais.

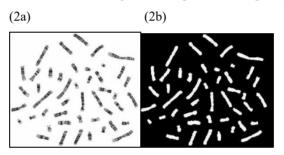
Devido ao tempo gasto para realizar todas essas etapas manualmente, e também devido à quantidade de metáfases que chegam aos laboratórios para serem cariotipadas, surge a necessidade de se automatizar esse processo. Apesar da demanda crescente, ainda é pequena a oferta de sistemas automáticos que auxiliem o trabalho dos geneticistas na coleta de dados e geração do cariótipo. O processo automático de identificação de cromossomos geralmente envolve as seguintes etapas:

- a) etapa de treinamento do sistema para que seja possível realizar a identificação de cromossomos, é preciso que se tenha descrito um padrão de suas características morfológicas. Esse padrão geralmente é obtido a partir de bases de imagens contendo centenas de imagens de cada tipo de cromossomo humano (do 1 ao 22, X e Y), e a partir daí, o sistema é treinado, geralmente por meio de algoritmos de inteligência artificial ou de métodos estatísticos, de maneira a tornar possível a identificação dos cromossomos de novas imagens.
- b) etapa de aquisição Da mesma forma que o processo manual, a imagem da metáfase é obtida por meio de uma câmera acoplada a um microscópio óptico com um aumento de 1000x;
- c) etapa de pré-processamento Essa etapa tem como foco principal a preparação da imagem para a etapa de segmentação, tornando possíveis as operações subsequentes, a fim de alcançar um resultado final esperado (ou pelo menos próximo dele). Essa etapa envolve a utilização de diversos filtros, tal como suavização, detecção de bordas, realce de contraste, entre outros registros;

- d) etapa de segmentação Etapa em que se extrai e se gera uma nova imagem para cada cromossomo da metáfase. Entre a etapa de segmentação e identificação ainda pode ocorrer uma nova etapa de préprocessamento, com a aplicação de novos filtros adequados, que auxiliem o processo de identificação e classificação;
- e) etapa de identificação e classificação Com as imagens segmentadas e preparadas após a etapa de segmentação, é iniciado o processo de identificação, utilizando-se alguma técnica/algoritmo desenvolvido, com base nos padrões obtidos na etapa 1.

No processo automático de identificação, a etapa de segmentação geralmente envolve algoritmos de detecção de bordas e de ajuste de contraste, de maneira a delimitar informações referentes ao que representa e ao que não representa um cromossomo na imagem (Figura 2). Já a etapa de identificação dos cromossomos, da mesma forma que o processo manual, envolve a extração de uma série de informações da imagem de forma a auxiliar e tornar possível a realização dessa tarefa, tais como o tamanho do cromossomo, a posição do centrômero, o padrão de bandas, entre outras informações [5]. Essas informações extraídas são então comparadas com as informações obtidas na base de imagens de treinamento, obtendo-se assim a classificação de cada cromossomo.

Figura 2: Exemplo de segmentação (b) de uma metáfase (a), buscando-se definir o que representa um cromossomo (cor branca) e o que não o representa (cor preta)



Fonte: elaborado pelos autores com dados da pesquisa.

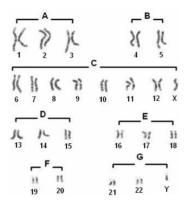
2.2 Características morfológicas dos cromossomos

Os cromossomos podem ser definidos como uma longa cadeia de DNA contendo vários genes, e essa cadeia está associada a diversas proteínas. Os cromossomos metafásicos apresentam uma morfologia característica e são constituídos por dois filamentos denominados cromátides, quesão unidos pelo centrômero (região mais condensada do cromossomo, ou seja, a região em que está concentrado o maior número de genes). Nos extremos dos braços estão os telômeros (região constituída por repetitivas sequências de DNA nos extremos dos cromossomos que os protegem contra degradação, recombinação e translocação) [2]. Os cromossomos possuem diversas características morfológicas que os diferem uns dos outros, permitindo assim que possam ser classificados. Dentre as principais características, podemos citar o tamanho, a posição do centrômero, o padrão de bandas e diversas outras características adicionais.

O centrômero divide as cromátides do cromossomo em dois braços, dos quais um, em geral, é mais longo do que o outro. O braço curto frequentemente é identificado pela letra p e o braço longo pela letra q. A divisão dos cromossomos pelo centrômero define uma característica importante que é muito utilizada na classificação de cada par, é a denominada Classificação de Denver.

Os cromossomos humanos, quando classificados com base no tamanho e na posição do centrômero, adotam um esquema estipulado, em 1960, em um congresso de citogeneticistas na cidade de Denver, daí a denominação Classificação de Denver. Nessa classificação, os cromossomos são divididos em sete grupos identificados por letras, de A a G, em ordem decrescente de tamanho, tal como é mostrado na Figura 3.

Figura 3: Classificação dos cromossomos de acordo com a Classificação de Denver



Fonte: elaborado pelos autores com dados da pesquisa.

O centrômero, conforme dito anteriormente, é a região que divide o cromossomo em braço curto e braço longo, sendo ela uma característica importante no processo de identificação dos cromossomos, pois a partir da posição do centrômero, os cromossomos podem ser classificados em três grupos: metacêntrico (centrômero na região central), submetacêntrico (longe da região central, com um braço maior que o outro) ou acrocêntrico (centrômero localizado em um dos extremos do cromossomo). A Figura 4 traz exemplos de imagens de cromossomos de cada tipo, obtidas na base de imagens utilizadas neste trabalho, com a região do centrômero marcada.

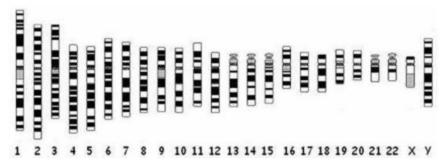
Figura 4: Exemplos de cromossomos metacêntricos (a), submetacêntricos (b) e acrocêntricos (c)



Fonte: elaborado pelos autores com dados da pesquisa.

No processo de cariotipagem, além do tamanho e da posição do centrômero, os cromossomos também possuem outra característica importante, que é o seu padrão de bandas. O padrão de bandas de um cromossomo são as faixas claras e escuras exibidas ao longo de seu eixo longitudinal, tal como é apresentado na Figura 5, na forma de ideogramas. A distribuição dessas faixas é diferente e constante em cada um dos cromossomos, o que facilita bastante a sua identificação [2].

Figura 5: Padrão de bandas do cariótipo humano na forma de ideogramas



Além de informações a respeito do comprimento, da posição do centrômero e do padrão de bandas de um cromossomo, os trabalhos de [5] [6] [7] destacam a importância de se utilizar outras características adicionais no processo de identificação automática. Algumas dessas características foram modeladas no XML resultante do trabalho proposto nesse artigo, tais como o perímetro do cromossomo, sua área e a média entre a área e o comprimento do cromossomo.

3 Bases de dados sobre cromossomos

Com o objetivo de desenvolver sistemas automatizados de identificação e classificação de cromossomos humanos a partir de imagens de metáfases, diferentes técnicas foram pesquisadas, desenvolvidas e publicadas em estudos anteriores. Dentre as técnicas, pode-se citar a utilização de modelos estatísticos [5] [8], redes neurais artificiais [9] [10] [11] [12], algoritmos de transporte [13], aprendizado incremental [14], lógica *fuzzy* [15], SVM (*support vector machine* – máquina de vetores de suporte) [16], entre outros. Dentre esses estudos, as técnicas que mais obtiveram sucesso nos resultados são as que envolvem algoritmos estatísticos ou redes neurais artificiais [12]. Na maior parte desses estudos, uma base de dados pública é utilizada como forma de treinamento para o sistema. As bases de dados mais utilizadas são as bases de Copenhagen [6] [17] [18], Edimburgo [6] [19] e Filadélfia [6] [17]. Essas bases de dados são constituídas por uma série de arquivos contendo informações na forma de *strings*, a respeito do padrão de bandas, do tamanho, do tamanho dos braços bem como a posição do centrômero de milhares de cromossomos. As bases de Edimburgo e Copenhagem são baseadas em imagens obtidas de células sanguíneas periféricas, enquanto que a base de Filadélfia é baseada em imagens obtidas de células extraídas do vilo coriônico [20].

Devido ao fato de essas bases de dados serem representadas e constituídas de um grande número de arquivos contendo uma enorme quantidade de informações não estruturadas, o acesso a essas informações e a sua alimentação com dados atualizados torna-se um processo bastante dispendioso. Além disso, devido ao fato de não existir um padrão sobre como essas informações estão organizadas, é necessário ter um conhecimento prévio de como esses arquivos estão dispostos bem como sobre as informações contidas. Por fim, algumas informações utilizadas no processo de identificação automática, tais como o perímetro e a área do cromossomo, não são apresentadas nessas bases de dados. A Figura 6 apresenta como estão dispostos os arquivos da base de dados de Copenhagem, e a Figura 7 apresenta o conteúdo de um dos arquivos, observando-se a não estruturação das informações e a quantidade excessiva de arquivos.

Figura 6: Disposição dos arquivos da base de Copenhagen



Fonte: elaborado pelos autores com dados da pesquisa.

Figura 7: Conteúdo de um arquivo da base de Copenhagen

```
3843 84 10 55 18
                       A=B===a==A==a==D==d====D==d===
                                                          =C==b=
                                                                 ==A==
7231 158 10 55 20 / A--B-a-C-a-A-c--D--d-----C--b--B---d--A-
787 15 10 55 18 / A-B-a-A--a-B--b--D--e---A--a-A--a-A--a-a-A
2459 60 10 54 19
                       A=B=aB==a=A==a==C==c==C===d=====C==b==
3290 21 10 54 19 /
                       A==B==a==A==a===B==b==D====c====B==b==A==c===A=a==a
5591 122 10 54 17 /
6359 139 10 54 18 /
                       A=A=a==A==a===B=
                       A===C==a==A==a==A==a==A==a==A===a==A==a=a
6399 140 10 54 20
                       A===B=a=B==a==A=aA=a==C===d===C==C===B==C
747 13 10 54 19 / A=
                     =B==a==A==a==A===a===D====d====C===b==A===c===A=a=a
792 15 10 54 18 / A==B===a=Aa===C====b===C====B===b===A==c===A=a=a
1785 66 10 53 18 /
4447 97 10 53 18 /
                       A-Aa-Aa-A-a----C--b---D--d-----A--a--B---c---A-a-a
                       AA--a--B--a--B--c--E--e---A---a--C--c---Aa-
                       A=B===a==A==a=A==a==D===c==B==
4634 101 10 53 18 /
                                                            ==b==A==c=
                                                                        =AaAa=a
     119 10 53 18
                       A=A==a=A===a===B==a===D==d=====A===a==A===b===
                                        =Δ=
```

Fonte: elaborado pelos autores com dados da pesquisa.

4 Metodologia

Para o desenvolvimento deste trabalho, foi utilizada a base de imagens disponibilizada pelo Laboratório de Imagem Biomédica [21]. A escolha dessa base de imagens é por ela apresentar arquivos de imagens de cromossomos já segmentados e com suas devidas classificações. Essa base de imagens é constituída de imagens de 119 metáfases que foram manualmente segmentadas e classificadas por especialistas, sendo ao total 5.474 imagens divididas em 119 pastas, cada pasta contendo imagens dos 46 cromossomos já separados. A partir dessa base de imagens, foram desenvolvidos algoritmos de processamento de imagens para a extração das principais características morfológicas desses cromossomos, conforme apresentado em trabalhos desenvolvidos anteriormente [22] [23]. Portanto, para cada uma das 46 imagens das 119 metáfases são extraídas as seguintes características morfológicas:

- Comprimento
- Área
- Perímetro
- Padrão de bandas
- Número de bandas
- Classificação do centrômero
- Posição do centrômero

O objetivo do XML é servir a dois propósitos: uma base de informações de amostras individuais de cromossomos e também como um modelo que descreve essas características para cada tipo de cromossomo, do 1 ao 22, e dos cromossomos sexuais X e Y. Para as informações individuais de cada cromossomo, a tarefa se torna mais simples, uma vez que é preciso simplesmente definir a estrutura dessas informações no XML Schema. Porém, para que se tenha um modelo dessas informações para cada um dos tipos de cromossomo, é necessário calcular uma média dessas características obtidas de cada imagem, pois a base de imagens contém 238 amostras de cada tipo de cromossomo.

O comprimento, a área e o perímetro de um cromossomo são representados por um valor real entre 0 e 1, sendo que esses valores são proporcionais ao maior valor obtido para cada uma dessas informações dentro de uma metáfase. Por exemplo, se em uma metáfase um dos cromossomos do tipo 1 for o cromossomo de maior comprimento dentre os 46, lhe é determinado comprimento de valor 1, e então todos os outros cromossomos da mesma metáfase terão seu comprimento definido com um valor proporcional em relação a ele. Ou seja, se o cromossomo 4 for a metade do comprimento do cromossomo 1 dessa metáfase, seu comprimento terá o valor de 0.5. O mesmo vale para a área e o perímetro. Por fim, é feita uma média para cada tipo de cromossomo entre todas as 119 metáfases.

O padrão de bandas de cada cromossomo é representador por um vetor de cem elementos do tipo inteiro, cujos valores de cada posição são determinados pela variação de níveis de cinza ao longo do eixo longitudinal do cromossomo, sendo que 0 representa a cor preta e 255 a cor branca. Como cada cromossomo possui um tamanho diferente, portanto, foi necessário realizar a padronização do tamanho desse vetor para todos os cromossomos, para que seja possível a realização de futuras comparações. Dessa forma, o tamanho do vetor que representa o padrão de bandas foi discretizado em cem elementos para todos os cromossomos. No XML, o vetor está representado por uma *string*, sendo que os valores de cada posição estão separados por vírgula.

O número de bandas é determinado pela quantidade de variações, a partir de um limiar, de níveis claros e escuros do padrão de bandas de um cromossomo. Ao final, é calculada a quantidade média do número de bandas para cada tipo de cromossomo.

Apesar de a classificação em relação ao centrômero ser sempre a mesma, a posição em que ele se encontra pode variar levemente entre dois cromossomos de um mesmo tipo. Dessa forma, o valor da posição do centrômero poderá variar entre 0 e 0.5, sendo que 0 representa um centrômero no extremo do cromossomo e 0.5 um centrômero no meio do cromossomo.

A partir do cálculo da média de cada característica, foi definido um XML Schema que descreve a estrutura do documento XML bem como um documento XML contendo essas informações.

5 Resultados

Nesta seção serão apresentados os resultados obtidos no desenvolvimento deste trabalho. O XML Schema, que descreve a estrutura do documento XML representando as informações morfológicas dos cromossomos, tanto individuais como dos modelos, é apresentado na Figura 8. A Figura 9 mostra uma parte do XML resultante, descrito conforme o XML Schema da Figura 6, apresentando as informações obtidas sobre o modelo de cromossomos do tipo 1 bem como uma amostra de um cromossomo do tipo 4.

Como se pode observar na Figura 8, tanto modelos quanto amostras seguem a mesma estrutura de informações, contendo o comprimento do cromossomo, sua área, o perímetro, o número de bandas encontradas, seu padrão de bandas, a classificação com relação ao centrômero e a posição do centrômero. Optou-se por não adicionar informações a respeito do tamanho dos braços, pois isso pode ser calculado a partir da posição do centrômero e sua classificação.

O arquivo XML da Figura 9 apresenta um exemplo de modelo de cromossomo 1, representado pelo elemento *modelo Cromossomo*. Conforme dito anteriormente, as informações do tipo *modelo Cromossomo* representam a média de todas as informações obtidas sobre o cromossomo daquele tipo específico (no caso da Figura 9, do cromossomo 1). Já o elemento *amostra Cromossomo* apresenta informações sobre uma amostra/imagem específica de um cromossomo do tipo 4.

Figura 8: XML Schema descrevendo a estrutura do documento XML

```
xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
xs:element name="cromossomos">
  <xs:complexType>
      <xs:sequence>
          <xs:element name="modeloCromossomo" type="tipoCromossomo" maxOccurs="unbounded"/>
           <xs:element name="amostraCromossomo" type="tipoCromossomo" maxOccurs="unbounded"/</pre>
/xs:element>
(xs:complexType name="tipoCromossomo">
  <xs:sequence>
      <xs:element name="comprimento" type="xs:decimal"/>
       <xs:element name="area" type="xs:decimal"/>
       <xs:element name="perimetro" type="xs:decimal"/>
       <xs:element name="numeroBandas" type="xs:integer"/>
       <xs:element name="padraoBandas" type="xs:string"/>
      <xs:element name="classificacaoCentromero" type="xs:string"/>
      <xs:element name="posicaoCentromero" type="xs:decimal"/>
   </xs:sequence>
  <xs:attribute type="xs:integer" name="tipo" use="required"/>
/xs:complexType>
```

Fonte: elaborado pelos autores com dados da pesquisa.

Figura 9: XML com informações sobre os cromossomos 1 e 4, descritas conforme o XML Schema da Figura 8

```
2xml version="1.0" encoding="utf-8"2
cromossomos>
 <modeloCromossomo tipo="1">
  <comprimento>0.94</comprimento>
  <area>0.96</area>
  <perimetro>0.91</perimetro>
  <numeroBandas>27</numeroBandas>
   <padraoBandas>226,224,213,207,202,198,195,193,191,189,187,185,183,
  180,177,174,171,167,163,160,155,152,148,144,140,135,133,130,128,126
  125,125,126,126,127,127,128,129,130,132,135,139,144,149,155,162,168,
  174,180,184,188,190,191,191,191,190,188,185,183,180,177,174,171,168,
  165,161,158,154,150,147,143,140,137,134,132,129,127,126,126,129,132
  136,139,142,145,146,147,148,149,151,152,154,156,157,160,163,170,179,
  192,212</padraoBandas>
  <classificacaoCentromero>Metacêntrico</classificacaoCentromero>
   <posicaoCentromero>0.48</posicaoCentromero>
 </modeloCromossomo>
 <amostraCromossomo tipo="4">
  <comprimento>0.74</comprimento>
  <area>0.81</area>
  <perimetro>0.72</perimetro>
  <numeroBandas>23</numeroBandas>
  <padraoBandas>218,218,206,195,184,174,164,155,147,140,134,129,125,
  121,118,116,115,114,113,113,113,114,115,115,116,117,117,118,117,117
  115,114,112,110,109,108,107,107,108,109,111,112,114,115,116,117,118
  117,117,117,116,115,115,115,115,114,114,114,114,113,113,112,112,111
  111,110,109,109,109,109,109,110,111,112,114,115,116,118,119,119,119
  119,119,118,118,117,118,118,119,121,124,127,132,137,146,156,167,180,
  195,210</padraoBandas>
  <classificacaoCentromero>Submetacêntrico</classificacaoCentromero>
  <posicaoCentromero>0.28</posicaoCentromero>
 </amostraCromossomo>
```

Fonte: elaborado pelos autores com dados da pesquisa.

6 Conclusão

Com o objetivo de organizar e estruturar as informações referentes ás características morfológicas de cromossomos humanos, de maneira a serem utilizadas em sistemas de identificação e classificação automática, desenvolveu-se, neste trabalho, uma simplificação de uma base de imagens, com as informações estruturadas por meio de XML Schema e organizadas no formato XML. A partir dos resultados obtidos, observa-se uma melhor organização das informações morfológicas dos cromossomos, comparada as bases de Copenhagen, Edimburgo e Filadélfia, previamente estudadas, que apresentavam as informações em uma série de arquivos. Ao extrair e estruturar as informações da base de imagens utilizando XML, tanto o acesso a essas informações quanto sua alimentação com novos dados, tornam-se processos menos dispendiosos. Além disso, ao centralizar, organizar e padronizar a forma com que os dados estão estruturados, possibilita-se seu entendimento e compreensão sem a necessidade de um conhecimento prévio sobre como estão dispostos.

Observa-se que a realização de tal estudo é de suma importância no que tange o desenvolvimento de sistemas automatizados, que serve de auxílio ao trabalho exaustivo do geneticista, pois o número de profissionais qualificados ainda é baixo para a grande demanda por profissionais capacitados na área de genética médica para a execução do cariótipo. A citogenética, como subárea da genética médica, surgiu com o intuito de pesquisar alterações que ocorrem em nível de DNA nas populações e estudar agentes mutagênicos. Uma das técnicas mais utilizadas e aplicáveis na realização desse trabalho é a análise do cariótipo. Dessa forma, a análise do cariótipo é essencial para o diagnóstico de doenças ligadas a alterações cromossômicas, e, portanto, a automatização desse processo e a representação organizada das informações referentes aos cromossomos é essencial, visando obter resultados mais confiáveis e em um espaço de tempo menor, permitindo também que o acesso e a alimentação dessas informações seja padronizada.

Este trabalho focou na representação das informações a respeito das características morfológicas de cromossomos humanos normais, visto que a base de imagens utilizada neste estudo contém somente imagens

desse tipo. Para a representação de anomalias cromossômicas morfológicas, necessita-se de uma base de dados que contenha imagens de anomalias cromossômicas, com uma classificação prévia das anomalias ocorrentes, além de que uma série de outras informações que precisariam ser extraídas das imagens. Em relação a trabalhos futuros, portanto, pretende-se realizar uma pesquisa sobre bases de imagens contendo anomalias morfológicas e então adicionar à modelagem do XML Schema uma estruturação dessas informações, visto que essa é uma das principais análises feitas pelos geneticistas.

Referências

- [1] BRUNONI, D. Estado atual do desenvolvimento dos serviços de genética médica no Brasil. *Revista Brasileira de Genética*, Ribeirão Preto, v. 20, Suppl. 1, p. 11-23, 1997.
- [2] BRUNONI, D. Aconselhamento genético. *Ciência & Saúde Coletiva*, Rio de Janeiro, v. 7, n. 1, p. 101-107, 2002.
- [3] SPERBERG-MCQUEEN, C. M.; THOMPSON, H. XML Schema. 2010. Disponível em: http://www.w3.org/XML/Schema. Acesso em: 28/06/2016.
- [4] NUSSBAUM, R. L.; MCINNES, R. R.; WILLARD, H. F. *Thompson e Thompson*: genética médica. 7. ed. Rio de Janeiro: Elsevier, 2008.
- [5] MORADI, M.; SETAREHDAN. New features for automatic classification of human chromosomes: A feasibility study. *Pattern Recognition Letters*, New York v. 27, p. 19-28, 2006.
- [6] PIPER, J.; GRANUM, E. On fully automatic feature measurement for banded chromosome classification. *Cytometry*, [S.I.*] v. 10, p. 242-255, 1989.
- [7] CONROY, J. M.; KOLDA, T. G.; O'LEARY, D. P. Chromosome identification using hidden markov models: comparison with neural networks, singular value decomposition, principal components analysis, and Fisher discriminant analysis. *Laboratory Investigation*, USA, v. 80, p. 1.629–1.641, 2000.
- [8] WU, Q.; LIU, Z.; CHEN, T.; XIONG, Z. K. R. C. Subspace-based prototyping and classification of chromosome images. *IEEE Trans Image Process*, League City, TX, Usa, v. 14, p. 1.277–1.287, 2005.
- [9] CHO, J.; RYU, S. Y.; WOO, S. H. A study for the hierarchical artificial neural network model for Giemsastained human chromosome classification. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY. *Proceedings*... San Francisco: IEEE 2004. p. 4.588-4.591.
- [10] GAGULA-PALALIC, S.; CAN, M. Human chromosome classification using Competitive Neural Network Teams (CNNT) and Nearest Neighbor. In: IEEE-EMBS INTERNATIONAL CONFERENCE ON BIOMEDICAL AND HEALTH INFORMATICS (BHI). *Proceedings*..., Valencia: IEEE 2014. p. 626-629.
- [11] OSKOUEI, B.; SHANBEHZADEH, J. Chromosome classification based on wavelet neural network. In: INTERNATIONAL CONFERENCE ON DIGITAL IMAGE COMPUTING: TECHNIQUES AND APPLICATIONS. *Proceedings*... Sydney: IEEE, 2010. p. 605-610.
- [12] WANG, X. et al. Automated classification of metaphase chromosomes: optimization of an adaptive computerized scheme. *Journal of Biomedical Informatics*, San Diego, USA, v. 42, p 22-31, 2009.
- [13] TSO, M. K. S., GRAHAM, J. The transportation algorithm as an aid to chromosome classification. *Pattern Recognition Letters*, [S.I.*], v. 1, p. 489–496, 1983.
- 14] JOSHI, P. et al. Efficient karyotyping of metaphase chromosomes using incremental learning. *IET Science, Measurement & Technology*, [S.I.*], v. 7, n. 5, p. 287-295, Sept. 2013.
- 15] KELLER, J. M. et al. A fuzzy logic rule-based system for chromosome recognition. In: IEEE SYMPOSIUM ON COMPUTER BASED MEDICAL SYSTEMS, 8. *Proceedings*... Lubbock, TX: IEEE, 1995. p. 125–132.

- [16] SARANYA, S.; LOGANATHAN, V.; RAMAPRABA, P. S. Efficient feature extraction and classification of chromosomes. INTERNATIONAL CONFERENCE ON INNOVATION INFORMATION IN COMPUTING TECHNOLOGIES. *Proceedings*... Chennai: IEEE, 2015. p. 1-7.
- [17] ZARDOSHTI-KERMANI, M.; AFSHORDI, A. Classification of chromosomes using higher-order neural networks. In: IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS. Proceedings... Perth, WA: IEEE, Nov./Dec. 1995. p. 2.587–2.591. v. 5.
- [18] LUNDSTEEN, C.; PHILIP, J.; GRANUM, E. Quantitative analysis of 6985 digitized trypsin G-banded human metaphase chromosomes. Clinical Genetics, [S.I.*], n. 18, p. 355–370, 1980.
- [19] LERNER, B. *Toward a completely automatic neural-network-based human chromosome analysis.* IEEE Transactions on Systems, Man, and Cybernetics Part B (Cybernetics), [S.I.*], v. 28, n. 4, p. 544–552, Aug. 1998.
- [20] KHMELINSKII, A.; VENTURA, R.; SANCHES, J. A novel metric for bone marrow cells chromosome pairing. IEEE Transactions on Biomedical Engineering, [S.I.*], v. 57, n. 6, June 2010.
- [21] POLETTI, E.; GRISAN, E.; RUGGERI, A. *Automatic classification of chromosomes in Q-band images*. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE-EMBS, 30. *Proceedings*... Vancouver: IEEE, 2008. p. 20-24.
- [22] KURTZ, G. C., LIBRELOTTO, G. R., PERLES, L. A.; SAGRILLO, M. *Identifição automática de cromossomos humanos*. BRAZILIAN SYMPOSIUM ON COMPUTER GRAPHICS AND IMAGE PROCESSING, 21. *Anais*... Campo Grande, MS: SBC, 2008. p. 33–36.
- [23] KURTZ, G. C. Metodologias para Detecção do Centrômero no Processo de Identificação de Cromossomos. Santa Maria: Universidade Federal de Santa Maria, 2011.