

Universidad del Bío Bío
Facultad de Ciencias Empresariales
Departamento de Sistemas de Información

Profesor Tutor: Tatiana Gutiérrez
Profesor Co-Tutor: Homero Urrutia



UNIVERSIDAD DEL BÍO-BÍO

**“DETECCIÓN Y CUANTIFICACIÓN DE BANDAS EN
IMÁGENES DE ELECTROFORESIS EN GEL DE
GRADIENTE DESNATURALIZANTE DGGE”**

Informe Final de Habilitación Profesional

Concepción, 27 de Febrero 2009

Carola Andrea Figueroa Flores
Ingeniería Civil en Informática

Resumen

El objetivo de este proyecto se enmarca en establecer una solución informática en un área específica de la Biología Molecular, siendo un área muy potente de desarrollo para el laboratorio de Biopelículas y Microbiología del Centro de Biotecnología de la Universidad de Concepción o Ceb-UDEC. Específicamente todo lo referente a la Biodiversidad Microbiana o diversidad de especies en el medio ambiente, para lo cual hacen uso de la técnica de la Electroforesis en Gel de Gradiente Desnaturalizante o DGGE.

La técnica del DGGE inicia con un número limitado de muestras, las cuales poseen en su interior secuencias de ADN (la que servirá como un identificador). Una vez que haya terminado la recolección de muestras, son colocadas cada una en tubos de ensayos diferentes, ya que estos serán usados para la amplificación del ADN a través de la Reacción en Cadena de la Polimerasa o PCR, y posteriormente serán depositados en la lámina de gel, la cual contiene en su interior numerosas columnas o carriles. En el momento que se hayan vaciado todos los tubos de ensayos sobre los carriles, esta lámina es llevada a la cámara de electroforesis en donde se aplicará electricidad, la que originará que estas secuencias vayan migrando hasta un cierto nivel, provocando en cada carril rayas o bandas. Cuando todo este proceso haya terminado (el cual tiene una duración de 3 días aproximadamente), la lámina resultante es fotografiada y visualizada por una cámara de luz ultra violeta, la cual se encuentra dentro de una cámara oscura.

Una vez obtenida la imagen DGGE en formato digital, comienza la etapa de su análisis o procesamiento, el cual es realizado a través del programa Quantity-One, sin embargo antes de que esto ocurra, el usuario debe asegurarse de que ésta sea del tipo TIF o Tagget Image File (Formato Digital de Fichero para Imagen), debido a que es el único formato digital que el programa puede procesar, en caso contrario se deberá posponer el análisis hasta que haya sido convertida a TIF, por medio de cualquier editor de imagen como el Photoshop o Corel Draw, sin embargo, este no es el único inconveniente que posee este programa, puesto que, la forma automática de detectar las bandas opera de manera deficiente y con un porcentaje considerable de error, debido a la falta de precisión en sus parámetros, para lo

cual se deberá realizar la detección de manera manual y en base a la experiencia del investigador en la mayoría de los casos. Además de este problema se encuentra, la falta de una alternativa relacionada con la cuantificación de las Bandas, la cual ayudaría considerablemente el estudio de la taxonomía de las especies, y por ende más material de estudio para nuevas investigaciones y/o publicaciones por parte del Ceb-UDEC.

Para entregar una solución eficiente a estos problemas encontrados, se ha realizado un constante análisis de la situación actual y de los requerimientos de los usuarios, además de la experiencia adquirida a través del entorno de trabajo en donde se han realizado numerosas pruebas de imágenes, las cuales en su mayoría han sido exitosas desde el punto de vista de la eficiencia, ya que se ha trabajado en disminuir las falencias encontradas por el Quantity-One. Es decir, el sistema construido trabaja sobre cualquier imagen, sin importar el tipo o el formato en que se encuentre, además de entregar una detección de bandas de manera automática más eficiente.

Cabe señalar además que el informe se ha dividido en cinco capítulos, definidos de la siguiente manera:

Capítulo I: Generalidades.

Este capítulo permite al lector introducirse al tema central del proyecto, entregando una descripción general sobre la solución propuesta y los objetivos que se deben de lograr.

Capítulo II: Aproximación a la Bioinformática

Este capítulo, permite al lector tener un mayor conocimiento sobre la Bioinformática y lo relacionado con las imágenes DGGE, dando énfasis en su funcionamiento y las tecnologías Bioinformáticas involucradas en su análisis.

Capítulo III: Análisis de Requerimientos.

El capítulo de Análisis de Requerimientos, tiene como objetivo principal, el poder identificar con mayor precisión las necesidades del laboratorio o área de estudio, a través de la identificación de los actores y la especificación de sus requerimientos, logrando sistema robusto y de mejor calidad

Capítulo IV: Trabajo desarrollado

En este capítulo se comentará el trabajo desarrollado concerniente a la solución propuesta, definiendo en él la definición de las actividades que realizará el sistema, los procesos para llevarlas a cabo y las metodologías de implementación.

INDICE GENERAL

Resumen.....	2
Índice General.....	5
Índice de Figuras.....	8
Índice de Tablas.....	11
Abreviaturas y/o Nomenclaturas.....	12
Capítulo I: Generalidades	
1.1. Presentación al tema.....	13
1.2. Justificación.....	16
1.3. Solución Propuesta.....	17
1.4.Objetivos del proyecto.....	17
1.4.1. Objetivo General.....	17
1.4.2. Objetivos Específicos.....	17
1.5. Descripción General del Sistema CYDB.....	18
1.5.1. Alcances y Límites del Sistema CYDB.....	19
1.5.2. Perspectiva del Producto.....	19
1.5.3. Características del Usuario.....	20
1.6.Beneficios del Sistema.....	21
1.7. Supuestos y Dependencias.....	21
Capítulo II: Aproximación a la Bioinformática	
2.1. Bioinformática su definición.....	22
2.2. Desarrollo de la Bioinformática.....	22
2.2.1. Principales áreas de investigación.....	24
2.2.2. Desarrollo de la Bioinformática en el Análisis de Imágenes.....	25
Capítulo III: Análisis de Requerimientos.	
3.1. Análisis de la situación actual del área de estudio.....	27
3.2. Identificación del Problema.....	31
3.2.1. Descripción del Problema de Detección e Identificación de Bandas.....	33
3.2.2. Descripción del Problema de Cuantificación de Bandas.....	36

3.3. Definición de requerimientos.....	37
3.3.1. Restricciones.....	37
3.3.2. Identificación de actores.....	38
3.4. Especificación de requerimientos.....	39
3.4.1. Requisitos de interfaces externas.....	39
3.4.2. Requisitos funcionales.....	40
3.4.3. Requisitos no funcionales.....	43
3.5. Componentes del Sistema.....	44
3.6. Diagrama de Paquetes y Arquitectura del Sistema.....	45
3.6.1. Diagrama de Paquetes.....	45
3.6.2. Arquitectura del Sistema.....	46
3.7. Diagramas y Casos de Usos.....	47
3.8. Análisis de riesgo.....	53
Capítulo IV: Desarrollo del Módulo CYDB.	
4.1. Definición de actividades del sistema CYDB.....	54
4.2. Descripción General de los sub-módulos de CYDB.....	56
4.3. Desarrollo del sub-módulo DIDB.....	57
4.4. Desarrollo del sub-módulo CDB.....	74
4.5. Tecnología utilizada en la Implementación.....	85
4.6. Implementación del Sistema CYDB.....	88
4.6.1 Pruebas en Diferentes Tipos de Imágenes DGGE.....	88
4.6.2 Aplicación del Módulo CDB.....	106
4.6.3. Generación de Reportes.....	109
4.6.4. Desarrollo de Interfaz.....	110
Conclusiones	112
Glosario	114
Referencias Bibliográficas	117
ANEXOS	118
Anexo A: Descripción de la Institución o Área de Estudio	118
Anexo B: Extracto del Paper: ” Band Detection in Random Amplified Polymorphism DNA	121

Images” del Dr. Luis Ruedas	
Anexo C: Extracto del informe final del Proyecto de Título sobre: “ Análisis de Imágenes, Metodología de Desarrollo DGGE y Tecnologías involucradas“	122
Anexo D: Funciones de Image Processing Toolbox	130
Anexo E: Funciones de Statics Processing Toolbox	131
Anexo F: Detalle de los resultados arrojados por el Módulo DIDB	132
Anexo G: Aplicación del Módulo CDB	144

INDICE DE FIGURAS

Figura 1a: Carril Seleccionado de la imagen DGGE a partir del Bromuro de Sodio	14
Figura 1b: Carril Seleccionado de la imagen DGGE a partir del Nitrato de Plata	14
Figura 2a: Carril de la figura 1a ampliado y girado. Con banda seleccionada	15
Figura 2b: Carril de la figura 1b ampliado y girado. Con banda seleccionada	15
Figura 3: Esquema General del Sistema CYDB.	19
Figura 4: Esquema General de la Disciplina de la Bioinformática	23
Figura 5: Migración de los Segmentos de ADN y Formación de Bandas	28
Figura 6: Metodología de Desarrollo de una DGGE con Nitrato de Plata	29
Figura 7: Metodología de Desarrollo de una DGGE con Bromuro de Sodio	30
Figura 8: Lámina DGGE provista de errores	31
Figura 9: Imagen DGGE con Bromuro de Sodio	32
Figura 10: Imagen DGGE con Nitrato de Plata	32
Figura 11: Arriba Carril original. Abajo: Bandas detectadas por el Q1	33
Figura 12: Imagen DGGE1_1	34
Figura 13: Procesamiento en Q1 de la imagen de la Figura 15	34
Figura 14: Problemas en la detección de Bandas	35
Figura 15: Aplicaciones referente a las Bandas	36
Figura 16: Usuarios finales	38
Figura 17: Comunicación entre Módulos	38
Figura 18: Esquema de Funcionamiento del Módulo CYDB	44
Figura 19: Diagrama de Paquetes	45
Figura 20: Diagrama de Paquetes en detalle	46
Figura 21: Arquitectura del Sistema	46
Figura 22: Diagrama Caso de Uso del Módulo SDC	47
Figura 23: Diagrama Caso de Uso del Sub-Módulo DIDB	49
Figura 24: Diagrama Caso de Uso del Sub-Módulo CDB.	51
Figura 25: Diagrama de Flujo de las Actividades del Sistema general	55
Figura 26: Diagrama de Flujo de los Pasos del Sub-Módulo DIDB	57

Figura 27: Imagen DGGE a Color	58
Figura 28: Imagen capturada a través de una cámara UV	59
Figura 29: Obtención del Negativo de una Imagen RGB..	60
Figura 30: Imagen obtenida desde el paso 2.	62
Figura 31: Identificación de los límites de cada carril (gelpos)	62
Figura 32: Extracción de los carriles	63
Figura 33: Carril Girado horizontalmente a partir de la imagen en A.	64
Figura 34: Histograma correspondiente a la suma de pixeles.	65
Figura 35: Demarcación de los principales ruidos presentados en el Histograma de suma de pixeles	66
Figura 36: Histograma de la suma de pixeles suavizado	67
Figura 37: Histograma de Normalización	68
Figura 38: Bandas Identificadas y Marcadas	69
Figura 39: Imagen DGGE de prueba	70
Figura 40: Carriles con las Bandas Identificadas	71
Figura 41: Concatenación de los Carriles a partir de la Figura 39.	72
Figura 42: Diagrama de Flujo del Módulo CDB	74
Figura 43: Carril obtenido para la detección de los límites de las bandas	76
Figura 44: Banda 1 Segmentada	78
Figura 45: Histograma de la Suma de Pixeles de la Banda	78
Figura 46: Curva de Suavizamiento del Histograma de Suma de Pixeles	79
Figura 47: Histograma de Normalización	80
Figura 48: Imagen DGGE [5] para la prueba 1.	89
Figura 49: Carriles extraídos de la imagen DGGE de la Prueba 1	90
Figura 50: Comparación de la detección de Bandas a través del módulo DIDB y Q1.	92
Figura 51: Gráfico comparativo entre el éxito en la detección en ambos programas.	92
Figura 52: Gráfico comparativo entre el error en la detección en ambos programas.	93
Figura 53: Imagen DGGE [6] para la prueba 2.	94
Figura 54: Carriles extraídos de la imagen DGGE de la Prueba 2.	98
Figura 55: Gráfico comparativo entre el éxito en la detección en ambos programas.	99

Figura 56: Gráfico comparativo entre el error en la detección en ambos programas.	99
Figura 57: Imagen DGGE, proporcionada por el laboratorio del Ceb-UDEC, para la prueba 3.	100
Figura 58: Carriles extraídos de la Imagen DGGE, de la prueba 3.	100
Figura 59: Gráfico comparativo entre el éxito en la detección en ambos programas.	103
Figura 60: Gráfico comparativo entre el error en la detección en ambos programas.	103
Figura 61: Gráfico comparativo sobre el éxito en la detección en las 3 pruebas.	104
Figura 62: Gráfico comparativo sobre el error en la detección en las 3 pruebas.	105
Figura 63: Imagen DGGE para la prueba 1.	106
Figura 64: Demostración de las etiquetas html dentro del lenguaje Matlab.	110
Figura 65: Plataforma del Sistema CYDB.	111
Figura 66: Plataforma del Sistema CYDB.	111
Figura 67: Organigrama del Ceb-UDEC	119
Figura 68: Imagen a escala de grises.	122
Figura 69: Imagen a color (RGB).	122
Figura 70: Imagen Vectorial de un círculo.	125
Figura 71: Imagen en Mapa de Bit de un círculo.	125
Figura 72: Ejemplo de Histogramas de frecuencia de intensidades en escala de grises.	125

INDICE DE TABLAS

Tabla 1: Límites de Cada Banda	80
Tabla 2: Cálculo de los Estimadores Estadísticos.	85
Tabla 3: Valores de la sumatoria de cada columna del eje X de la Banda 1.	87
Tabla 4: Resumen de los Valores de Cuantificación.	87
Tabla 5: Resumen de la Participación de Cada Banda en el Carril.	88
Tabla 6: Cuadro comparativo de las Bandas encontradas en la Prueba 1.	92
Tabla 7: Cuadro comparativo de las Bandas encontradas en la Prueba 2.	92
Tabla 8: Cuadro comparativo de las Bandas encontradas en la Prueba 3.	96
Tabla 9: Cuadro comparativo con las Bandas Detectadas de manera exitosa y con error	97
Tabla 10: Cuadro comparativo con las Probabilidades de Error y Éxito en cada programa, para la prueba 2.	98
Tabla 11: Cuadro comparativo de las Bandas encontradas en la Prueba 3.	102
Tabla 12: Cuadro comparativo con las Bandas Detectadas de manera exitosa y con error.	103
Tabla 13: Cuadro comparativo con las probabilidades de éxito y de error en la detección de bandas en ambas aplicaciones.	103
Tabla 13: Cuadro resumen con los resultados de las tres pruebas.	105
Tabla 14: Participación y Cuantificación de las Bandas	110
Tabla 15: Cuadro Resumen de la Cuantificación en la Imagen DGGE 1.	110

Listado de Nomenclatura y/o Abreviaturas

1. **PCR:** Reacción en cadena de la Polimerasa
2. **DGGE:** Electroforesis en Gel de Gradiente Desnaturalizante.
3. **PM:** Peso Molecular.
4. **ADN:** Ácido desoxirribonucleico.
5. **Tm:** Temperatura de fusión.
6. **ARN:** Ácido ribonucleico.
7. **Ceb-UDEC:** Centro de Biotecnología de la Universidad de Concepción.
8. **Q1:** Abreviación para el software Quantity-One utilizado por Ceb-UDEC.
9. **CYDB:** Módulo Cuantificación y Detección de Bandas.
10. **DIDB:** Sub- Módulo Detección e Identificación de Bandas.
11. **SDC:** Sub-Módulo Sistema de Carriles.
12. **CDB:** Sub-Módulo Cuantificación de Bandas.
13. **UV:** Cámara Ultra Violeta
14. **GDR:** Generador de Reportes.
15. **SCOB:** Sistema de Corrección de Bandas.

Capítulo I:

Generalidades

Para facilitar la comprensión del lector sobre el tema central del proyecto, se ha desarrollado este capítulo, el cual abordará de manera general las características del sistema a construir, su propósito, beneficios, límites y los objetivos que se deben de alcanzar. Además de una breve descripción sobre la solución propuesta, relacionada con el Análisis de Imágenes, la cual es una de las principales áreas de la Bioinformática y la plataforma a utilizar en el desarrollo del sistema.

1.1. Presentación al tema.

El Análisis de Imágenes, antes era muy común que se llevara a cabo de forma manual o visual, lo que incurría a equivocaciones, debido a la falta de precisión que existía en ello. Sin embargo, para contrarrestar esta situación, surge el Análisis de Imágenes digital a través de aplicaciones computacionales e informáticas, permitiendo la automatización del procedimiento manual, además de garantizar calidad en los resultados esperados. Entre estas aplicaciones computacionales, se encuentran las relacionadas con la Bioinformática, la cual depende de un conjunto de conocimientos como: Inteligencia Artificial, Lógica Aplicada, Estadística y los Fundamentos Principales de la Informática, permitiendo entregar un mejor apoyo a biólogos, químicos o bioquímicos, en la realización de sus proyectos y/o publicaciones, como por ejemplo el estudio relacionado con el área de la Biodiversidad, o la diversidad de especies que existen en el ecosistema.

El Centro de Biotecnología de la Universidad de Concepción o Ceb-UDEC¹, cuenta con un listado de metodologías y técnicas, que permiten enriquecer el conocimiento sobre esta materia. Siendo una de ellas, la técnica de Electroforesis en Gel de Gradiente Desnaturalizante o DGGE², técnica que permite trabajar a partir de un número determinado

1 Ver Anexo A: “Descripción de la Institución o Área de Estudio”

2 Ver Glosario.

de muestras de ADN, las que son depositadas en una lámina de gel (hecha de agarosa) compuesta en su interior por columnas o carriles, como se puede apreciar en la Figura 1a y 1b, en donde cada carril contendrá a una de estas muestras, las que comenzarán a migrar a un nivel de temperatura constante igual a 60°C, dando como resultado a diversas marcas o bandas como se observa en la Figura 1a y 1b.

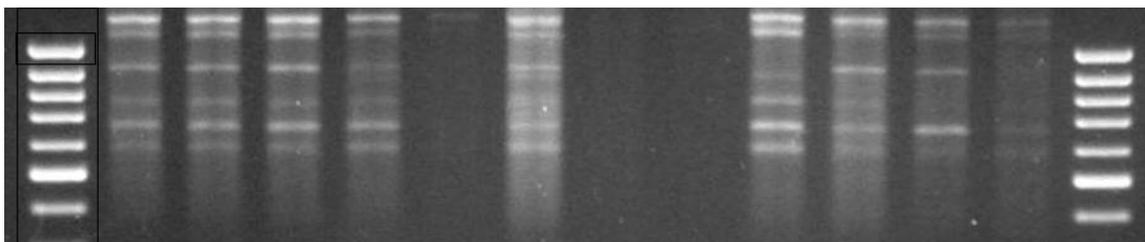


Figura 1a: Carril Seleccionado de la imagen DGGE a partir del Bromuro de Sodio³

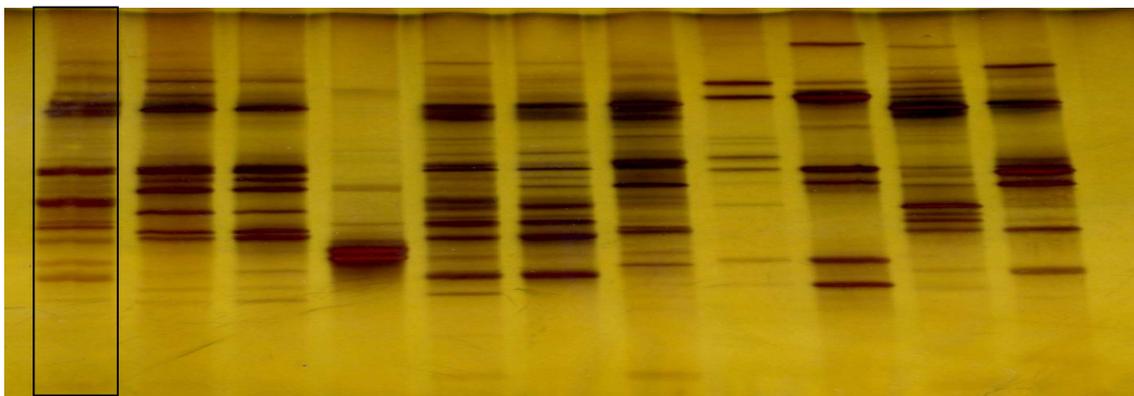


Figura 1b: Carril Seleccionado de la imagen DGGE a partir del Nitrato de Plata⁴

Las cuales simbolizan las Secuencias de ADN⁵ de las distintas especies que se pudieran encontrar en dicha muestra. Estas bandas se pueden visualizar de mejor manera en la Figura 2a y 2b, en donde muestra al primer carril de la Figura 1a y 1b respectivamente, con un tamaño mayor al original⁶ y de forma horizontal

3 y 4 Pigmentación que se le agrega a la imagen para obtener ese color

5 Ver Glosario

6 Dimensión original 47 x 192 pixeles.

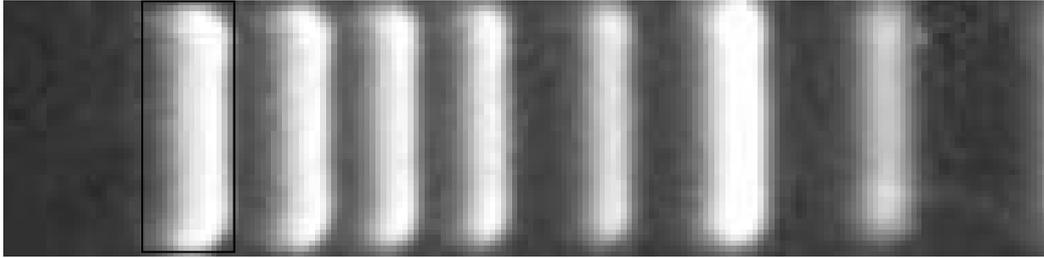


Figura 2a: Carril de la figura 1a ampliado y girado. Con banda seleccionada



Figura 2b: Carril de la figura 1b ampliado y girado. Con banda seleccionada

Una vez terminada la lámina de gel, es fotografiada dando origen a una imagen en formato digital, con el propósito de que el software utilizado por el laboratorio, pueda realizar la etapa referente al análisis y procesamiento de la imagen. Este software llamado Quantity One⁷ o Q1, ha sido hasta el momento un excelente apoyo, en las investigaciones realizadas por el laboratorio, sin embargo, no está exento a errores, ya que una de sus deficiencias radica en la detección automática de las Bandas y el estar limitado a trabajar sólo en imágenes TIF. Ésta situación por ende hace que la labor sea más engorrosa de lo normal, ya que en primer lugar el usuario debe verificar el formato de la imagen y tener conocimiento previo de ello. Sumándose a ello el hecho, de que al terminar la ejecución el usuario en base a su experiencia como investigador, deberá agregar o descartar aquellas bandas que fueron detectadas erróneamente de forma manual, desconfiando así en un 95% de los resultados arrojados de manera automática, como se puede apreciar en las pruebas realizadas en el Capítulo V de este documento.

7 Versión 4.2.3 para plataformas Windows.

Se puede determinar entonces que los problemas a enfrentar en este proyecto son:

1. La falta de sensibilidad y precisión, en determinar las bandas dentro de una imagen DGGE.
2. La inexistencia de aplicaciones que permitan llevar a cabo la cuantificación de las bandas.

1.2. Justificación

El continuo desarrollo de aplicaciones Informáticas ha ido aumentando cada día más en cuanto a su funcionalidad y usabilidad, en distintas áreas tales como: la educación, la investigación, la tecnología y la ciencia; en donde abre una amplia gama de posibilidades en donde se podría desempeñar, siendo la de interés las ciencias Biológicas, específicamente en el estudio de la Biodiversidad Microbiana. En donde el protagonismo recae en la información proporcionada por el ADN, para las diversas técnicas que son usadas para el reconocimiento y similitud de especies u organismos, siendo una de ellas la técnica del Análisis de Imágenes DGGE.

Este tipo de análisis, como se ha señalado es de gran importancia es por ello que el laboratorio del Ceb-UDEC ha solicitado mejorar dicho procesamiento, a través de la construcción de un sistema, con el fin de contar con una alternativa de mejor calidad que la que existe en el presente, la cual reducirá considerablemente la tasa de error estimada entre un 75% en la detección automática de las Bandas, mejorando de esta manera la calidad de los resultados esperados.

1.3. Solución Propuesta.

Una vez conocida la problemática actual del laboratorio del Ceb-UDEC, se podrá describir la solución propuesta, la cual se fundamenta en el Análisis de Imágenes⁸, el uso de la Estadísticas, y todo ello aplicado en MATLAB, ya que será la herramienta de trabajo, para el desarrollo del módulo de Detección, Identificación y Cuantificación de Bandas en imágenes DGGE.

1.4. Objetivos del proyecto.

1.4.1. Objetivo General.

Construir un sistema que permita a partir de imágenes DGGE en formato digital, identificar, detectar, cuantificar las bandas y la generación de reportes sobre la información que se encuentra en ella, para determinar la estructura taxonómica de los organismos en estudio por parte de los biólogos del laboratorio del Ceb-UDEC.

1.4.2. Objetivos Específicos.

Con el propósito de lograr el objetivo antes definido, se deberá lograr los siguientes objetivos específicos, los cuales fueron divididos en base a la etapa de Proyecto de Título y Habilitación Profesional:

Proyecto de Título

1. Se logró capturar adecuadamente los requerimientos de los principales usuarios del sistema, los cuales permitieron construir una aplicación que refleje de manera correcta las necesidades y exigencias de cada uno de ellos.

⁸ Ver Anexo C: “Extracto del Informe final del Proyecto de Título sobre el Análisis de Imágenes”

2. Se debió presenciar diferentes tipos de experimentos realizados sobre las imágenes DGGE, para conocer su comportamiento y características.
3. Se realizaron múltiples investigaciones sobre el análisis de señales, procesamiento de imágenes y técnicas de inteligencia artificial. Debido a que estos fueron las bases para la construcción del sistema.
4. Se debió entregar a fines del semestre pasado un informe que resumiera, todo el trabajo realizado durante la etapa del proyecto de título y los avances que se han hecho con respecto a la habilitación profesional.
5. Se debió realizar además, una aplicación en MATLAB, para el procesamiento de imágenes.

Habilitación Profesional

1. Afinar y corregir los requerimientos capturados en la etapa de Proyecto de Título.
2. Depurar los sub-módulos DIDB y CDB.
3. Realizar diferentes pruebas en distintos tipos de imágenes DGGE y en cualquier formato.
4. Realizar una aplicación que permita generar algún reporte de los resultados de los sub- módulos.
5. Depurar los sub-módulos DIDB y CDB.
6. Entregar el sistema terminado con su documentación.

1.5. Descripción General del Sistema DGGE

El sistema a construir será un módulo integrable al Sistema General DGGE, como puede apreciarse en la Figura 3, el cual deberá de contener a este módulo CYDB con sus sub-módulos, correspondiente a la Detección e Identificación de Bandas (DIDB), Extracción de Carriles (SDC) y el de Cuantificación de Bandas (CDB), respectivamente en conjunto con los demás, como por ejemplo el módulo de Sistema de Interfaz o SDI, el cual realiza la labor de integrar todos los módulos del Sistema General DGGE incluyendo al CYBD.

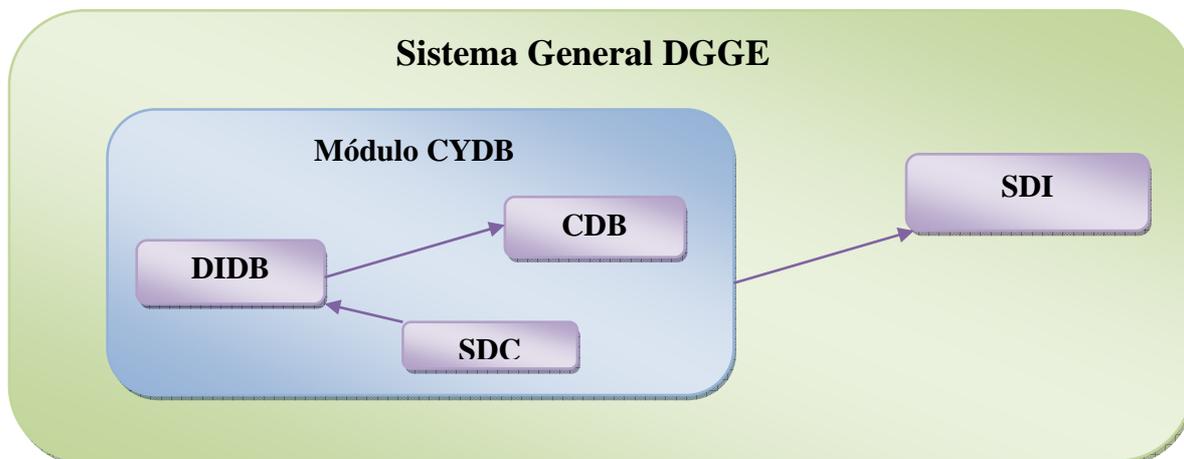


Figura 3: Esquema General del Sistema CYDB.

1.5.1. Alcances del Sistema CYDB.

El sistema a realizar tiene los siguientes alcances de desarrollo los cuales comprenden: la detección de bandas, su cuantificación y la generación de reportes asociada a la información proporcionada. Dejando de manera opcional el hecho de crear una interfaz de usuario, y fuera del alcance la corrección de Bandas y análisis posterior a la clasificación, ya que éstos son temas que se dejan propuestos para proyectos futuros.

1.5.2. Perspectiva del Producto.

Este tipo de sistema desde la perspectiva del producto, deberá ser capaz de diferenciarse de manera cualitativa y cuantitativa, respecto de la aplicación utilizada hasta la fecha por el área de estudio, que es el software Quantity-One, de manera que el sistema en construcción deberá de tener una propiedad única, para sacar de ello una ventaja comparativa con el resto de las aplicaciones, como el CMID, One Scan-IT o el Q1.

Con respecto a la perspectiva de los usuarios del producto, se puede mencionar que éstos deberán de quedar satisfechos con el producto una vez terminado, siendo de gran importancia el poder capturar de manera adecuada todos los requerimientos de los usuarios del sistema. Se puede agregar además que la calidad de la perspectiva del producto,

también está influenciada por la tecnologías de información a utilizar para la calidad de los datos, la cual implica que éstos al ser capturados, procesados, almacenados y entregados sean un fiel reflejo de la realidad que se desea tratar mediante este sistema. Esto supone que los datos no contengan errores, sean veraces y estén actualizados.

La calidad del producto significa por lo tanto aportar valor al usuario, esto es, ofrecer unas condiciones de uso del producto o servicio superiores a las que el usuario espera recibir y sin costo monetario.

1.5.3. Características del Usuario

Los usuarios de este sistema, se caracterizan por pertenecer al ámbito de la Biología, específicamente son investigadores que se encargan de estudiar todo lo relacionado con la Biodiversidad para lo cual se dedican a la extracción de muestras de material genético de un conjunto de organismo, que en su mayoría son bacterias como la E. Coli o Birreactores de Residuo, utilizados para el tratamiento de la contaminación ambiental.

Dentro de este tipo de usuarios, se encuentran los usuarios directos y los indirectos del sistema:

- *Directo*

Son los que utilizarán el sistema para el uso académico, como investigaciones y tareas escolares.

- *Indirecto*

En esta clasificación se encuentran los que se beneficiaran a partir de la información proporcionada por este sistema, la cual puede ser a través de publicaciones de los mismos usuarios directos.

Se puede agregar que el sistema no necesitará de administradores, por lo tanto no existirán restricciones de quién pueda usarlo, ya que la seguridad está a cargo del Ceb-UDEC de quién ingresa al establecimiento y específicamente al laboratorio o área de estudio.

1.6. Beneficios.

Los principales beneficios que se obtienen del sistema, es el hecho de poder leer y procesar cualquier tipo de imagen digital, es decir, que esta pueda estar en formato TIF, JPEG, GIF, PNG, debido a que el Q1 está restringido solo a la lectura de imágenes TIF, lo que conlleva a que el usuario deba de percatarse que tipo de imagen sea antes de usar el software. Desde el punto de vista funcional, el sistema CYBD ha demostrado poseer un nivel de precisión en la detección de Bandas mucho mayor que el sistema actual. Además se espera que sea un aporte en las nuevas investigaciones y publicaciones relacionadas con el área de la Microbiología o Biodiversidad, en cuanto a la calidad de los resultados obtenidos en la detección y por el hecho de contar con una técnica que sirva de alternativa al momento de realizar la cuantificación.

1.7. Supuestos y Dependencias.

En esta sección se describen los supuestos y dependencias, que pueden afectar la viabilidad de la aplicación a construir. Uno de estos inconvenientes, se pueden resumir de la siguiente manera:

- *La disponibilidad del sistema Hardware*
Que el sistema operativo sea compatible con las mínimas exigencias establecidas.
- *La disponibilidad de los otros sistemas o módulos y sus componentes*
Como esta aplicación será un componente más de un sistema mayor, deberá proporcionar a los demás módulos y éstos a él, de la información obtenida a partir de los datos ingresados, los cuales en este caso serían las imágenes digitales de DGGE.
- *La mínima capacidad técnica requerida*
El sistema al finalizar su construcción deberá de proporcionar un manual de usuario para su utilización.

Capítulo II:

Aproximación a la Bioinformática

Este capítulo se creó con el fin de entregar una información más detallada con respecto a la Bioinformática, debido a que el sistema se desarrolla en base a sus fundamentos en el área del Análisis de Imágenes DGGE. Además para informar al lector de las demás disciplinas en donde la Bioinformática se está desempeñando y la importancia en el desarrollo de las ciencias biológicas.

2.1. Bioinformática su definición

Disciplina científica dedicada a la investigación y desarrollo de herramientas útiles para llegar a entender el flujo de información desde los genes a las estructuras moleculares, a su función bioquímica, a su conducta biológica y, finalmente, a su influencia en las enfermedades y en la salud [1].

2.2. Desarrollo de la Bioinformática

La Bioinformática y la Biología Computacional involucran el uso de diversas técnicas, como: Matemática Aplicada, Informática, Estadística, Computación, Inteligencia Artificial, Química y Bioquímica, todo ello esquematizado en el diagrama de la Figura 4.

La Bioinformática a diferencia de la Biología Computacional, opera principalmente en la creación y desarrollo de algoritmos (genéticos en su mayoría), junto a técnicas Estadísticas, Computacionales y Análisis Biológico. En cambio la Biología Computacional, está enfocada sólo a la investigación a través de hipótesis, haciendo uso de computadoras, a través de datos experimentales y simulados, donde el objetivo principal es el descubrimiento y el avance del conocimiento biológico.

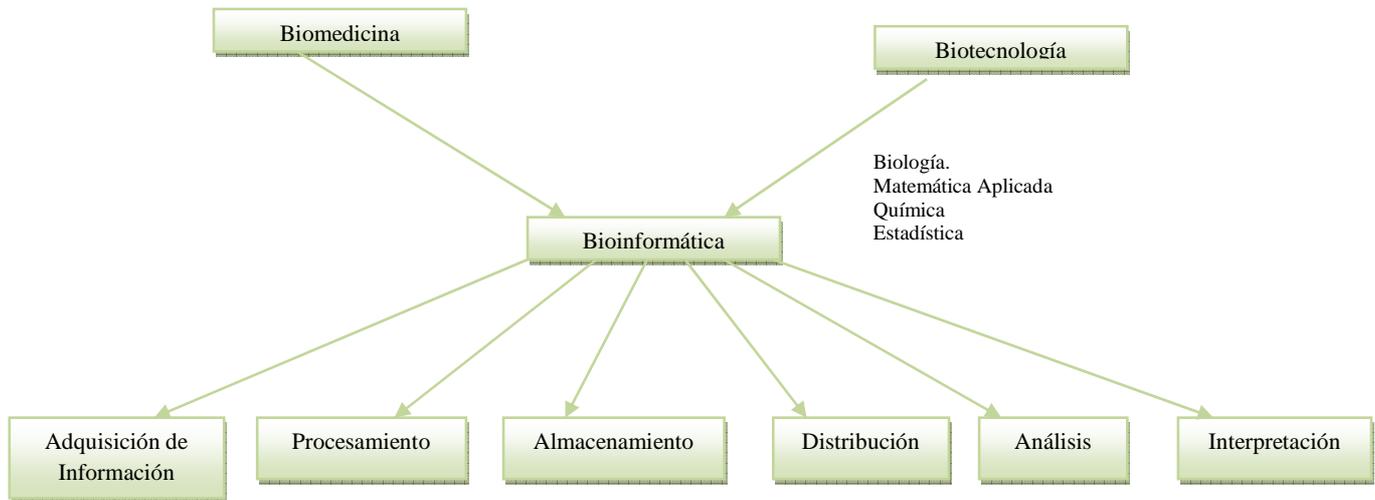


Figura 4: Esquema General de la Disciplina de la Bioinformática

Una de las características principales de esta disciplina se resume de la siguiente manera:

- Interdisciplinaria ya que facilita la colaboración de grupos relacionados con la Biología, Medicina e Informática.
- Formación de redes entre los grupos interdisciplinarios.
- Análisis de datos de proyectos biológicos de gran escala, como el Genoma y las Proteínas.
- Bases de datos disponibles online, como NCBI, UNIPROT, etc. [2].

Su objetivo principal se fundamenta de la siguiente manera:

- Profundizar en el entendimiento acerca de los organismos vivos y sus relaciones, partiendo desde el genoma que les codifica.
- Lograr la organización de la información con la creación y mantenimiento de bases y bancos de datos
- Utilizar la algoritmia, a través de algoritmos de búsqueda, alineamientos de secuencias.

- Realizar búsqueda de genes, por medio de la modelización y análisis estadístico; además de la programación dinámica.
- Lograr conocimientos en el área de proteína, con el estudio su estructura y la predicción de su funcionamiento.

2.2.1. Principales áreas de investigación

Una de las principales áreas de investigación en donde tiene campo la Bioinformática, se pueden resumir de la siguiente manera [3]

- **Análisis de Secuencias**
Permite identificar y analizar secuencias de proteínas que se encuentren relacionadas, de manera automatizada y precisa, sin que éstas sean idénticas o que provengan del mismo ancestro
- **Anotación de Genomas**
Permite caracterizar los genes o secuencias de ADN, para su localización con el fin de codificar una proteína.
- **Filoinformática**
Permite estudiar el origen ancestral de las especies, como el cambio que han sufrido a través del tiempo, en diferentes campos de esta área.
- **Medición de la Biodiversidad**
Con la ayuda de Base de Datos, se pueden recoger los nombres de las especies, con sus descripciones, información genética, estado y tamaños de las poblaciones, para un determinado ecosistema.
- **Análisis de la expresión genética**
Permite identificar los genes implicados en un micro arreglo de células, como por ejemplo las cancerosas con otras que no lo son, para determinar las transcripciones que son activadas o reprimidas en una población específica de esas células afectadas por el cáncer.
- **Análisis de Mutaciones en el Cáncer**

A través de la automatización se puede gestionar grandes volúmenes de secuencias y a partir de ellas crear nuevos algoritmos y software, para comparar los resultados obtenidos con las nuevas secuencias del genoma humano.

- **Comparación Genómica**

Permite el rastreo de los procesos evolutivos responsables de la diferencia entre dos genomas, a través de modelos y algoritmos matemáticos

- **Análisis de Imágenes**

Permite automatizar el proceso de observación visual de varias imágenes, por ejemplo la cuantificación de células.

- **Proteómica**

Permite el estudio de las interacciones entre las proteínas.

- **Bioinformática Estructural**

Permite estudiar la estructura y el plegamiento de las principales macromoléculas, como lo son los aminoácidos y proteínas, a través del desarrollo de algoritmos.

2.2.2. Desarrollo de la Bioinformática en el Análisis de Imágenes

El análisis de imágenes como se expuso en el punto anterior de este capítulo, es una de las tantas áreas en donde la Bioinformática ha podido desarrollarse cada día más (siendo además el área de interés para el proyecto). En la cual se han hecho uso de diversas tecnologías relacionadas con la computación, con el fin de automatizar completamente el procesamiento, la cuantificación y el análisis de grandes cantidades de imágenes biológicas, biomédicas, las cuales poseen un alto contenido de información, muy importante para quienes hacen uso de ella.

Estos nuevos sistemas de análisis de imagen agilizan la habilidad del observador para realizar sus investigaciones, mejorando con ello la precisión y la objetividad de los resultados arrojados. Un sistema de análisis totalmente desarrollado podría reemplazar completamente al observador. Aunque estos sistemas no son exclusivos del campo de las imágenes biomédicas, cada vez son más importantes tanto para el diagnóstico como para la investigación.

Algunos ejemplos de ello se resumen en:

- Cuantificación y localización subcelular con alta productividad y precisión.
- Análisis y visualización de imágenes clínicas.
- Determinación de patrones en el flujo del aire en tiempo real de la respiración pulmonar de animales vivos.
- Cuantificación del tamaño de la oclusión a través de imágenes en tiempo real, tanto por desarrollo como por recuperación, de lesiones arteriales.
- Realización de observaciones conductuales basadas en prolongadas grabaciones en vídeo de animales de laboratorio.
- Observaciones en infrarrojo para la determinación de la actividad metabólica

Capítulo III:

Análisis de Requerimientos

El capítulo de Análisis de Requerimientos, tiene como fin principal el describir de manera detallada lo que esperan los usuarios del sistema, incluyendo un análisis de la situación actual, permitiendo con ello el poder identificar los requerimientos tanto funcionales como no funcionales y los posibles riesgos a los cuales estará sometido el sistema CYDB.

La información en detalle asociada a la institución o área de estudio, se encuentra en el Anexo A.

3.1. Análisis de la situación actual del área de estudio.

El laboratorio de Biotecnología y Microbiología⁹ del Ceb-UDEC, tiene como propósito principal desarrollar investigaciones asociadas al tema de la Biodiversidad Microbiológica. Este tipo de investigación se origina con la toma de muestras, compuestas principalmente de Secuencias de Nucleótidos de ADN, las cuales son analizadas a través de la creación de Geles de Electroforesis en Gradiente Desnaturalizante o DGGE.

El motivo de la utilización del DGGE, es que permite obtener una mejor visualización del ADN, basado en la técnica de la Reacción en Cadena de la Polimerasa o PCR.

Una vez que se ha seleccionado el segmento de ADN y se haya aplicado el PCR, cada muestra es depositada en un Carril diferente dentro de la lámina de gel, la cual es depositada finalmente en el DCODE UNIVERSAL MUTATION, el cual se encarga de la Electroforesis y de que cada secuencia migre hasta un cierto nivel determinado por el GCclamp, formando así las marcas o bandas como se observa en la Figura 5.

⁹ Ver Anexo A: “Descripción de la Institución o Área de Estudio”

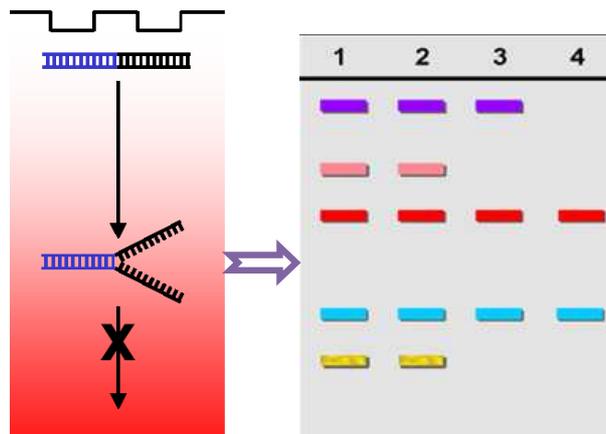


Figura 5: Migración de los Segmentos de ADN y Formación de Bandas

Todo este proceso descrito anteriormente puede resumirse en la Figura 6 y 7, respectivamente, en la cual la primera representa la obtención de una imagen con la ayuda del Nitrato de Plata, la que le da una tonalidad más café a las Bandas, y la segunda figura a la imagen efectuada a través del Bromuro de Sodio, que entrega una imagen en negro y blanco la cual solo puede ser visualizada a través de una cámara de UV. Todo este proceso tiene una duración estimada de 3 días en el mejor de los casos, lo que implica que cualquier error en la toma de muestras o en la obtención del ADN, retrasaría considerablemente cualquier experimento. Ya que al finalizar el DGGE, no se apreciaría en la lámina a ninguna Banda.

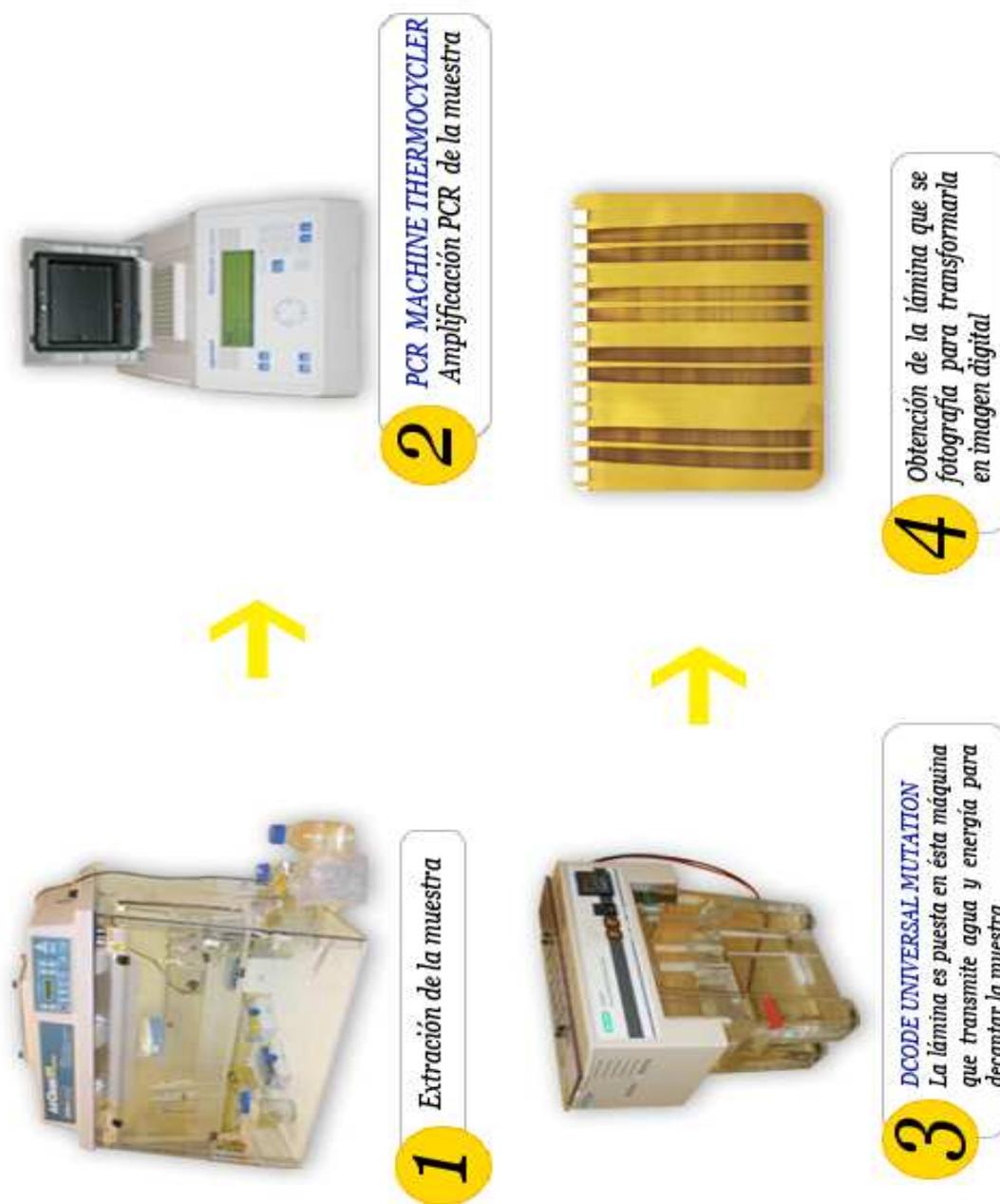


Figura 6: Metodología de Desarrollo de una DGGE con Nitrato de Plata¹⁰.

10 Diagrama a partir de fotografías realizadas en el laboratorio del Ceb-UDEC.



Figura 7: Metodología de Desarrollo de una DGGE a partir del uso del Bromuro de Sodio.

3.2. Identificación del Problema.

Una vez que se haya finalizado el procedimiento demostrado en las Figuras 6 y 7, la lámina resultante es observada detenidamente por la cámara UV¹¹, con el fin de poder decidir si es descartada o no, un ejemplo de ello es el caso que se observa en la Figura 8, en donde la lámina queda inutilizable para un posterior análisis.

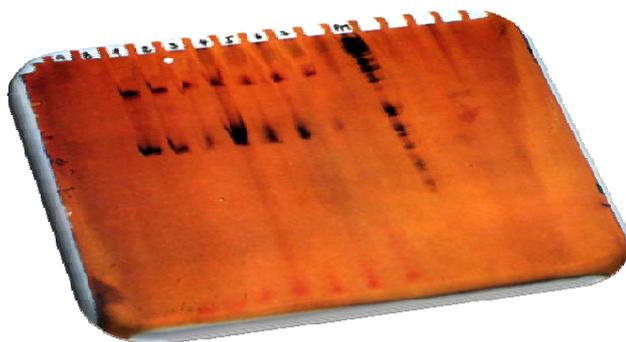


Figura 8. Lámina DGGE provista de errores

A pesar de que este problema no corresponde al desarrollo de la solución propuesta, es un factor que ha afectado en cierto grado el avance de las pruebas del Módulo DIDB, ya que la mayoría de las pruebas han sido hechas en imágenes descargadas desde la web, debido a problemas técnicos en la confección de geles DGGE, por parte de los investigadores del área de estudio.

No obstante si la lámina resultante hubiera sido la esperada, es decir, hubiera tenido las características que se esperan, con respecto a la visualización de bandas en los carriles como se ve en la Figura 9 y 0, las cuales corresponden a la utilización del Bromuro de Sodio y Nitrato de Plata respectivamente, siendo la primera la más utilizada en el presente debido a que el sistema Q1, opera de mejor manera sobre imágenes en escala de grises.

11 En el caso de la Figura 9.

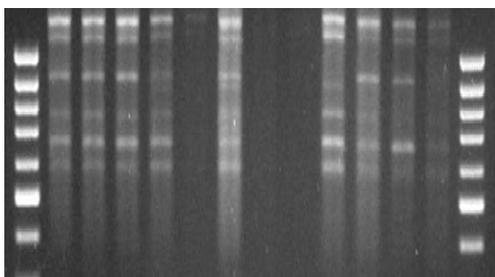


Figura 9 Imagen DGGE con Bromuro de Sodio

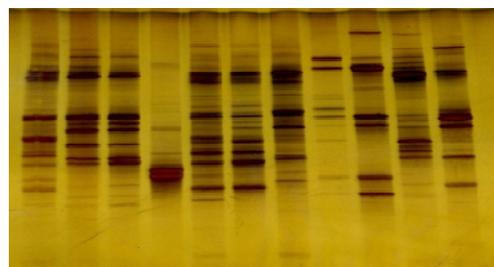


Figura 10 Imagen DGGE con Nitrato de Plata

A pesar de que esta sea una deficiencia del software, es decir, que sea capaz solo de leer imágenes en escala de grises, no es la única que posee, ya que el único formato que permite procesar corresponde al TIF, siendo un inconveniente si un usuario quiere analizar una imagen con distinto formato al requerido, deberá recurrir a otro programa como el Corel Draw o Photoshop¹² para realizar el cambio al formato, lo que conllevaría a atrasar más los resultados relacionados con la Biodiversidad. A todo esto se agrega además la falta de precisión en la Identificación de las Bandas de manera automática, la cual en el 90% de los casos el usuario debe de corregir manualmente lo que a su parecer corresponde o no a Banda, lo que se transformaría en una labor tediosa para él. Sumándose a ello, el hecho de no contar con alguna aplicación que permita cuantificarlas, ya que se da por hecho que el DGGE es una técnica Cualitativa, para lo cual el laboratorio ha solicitado que esta situación se pueda revertir, es decir, poder obtener una aplicación que permita cuantificar estas bandas de algún modo, con el fin de determinar de mejor manera la estructura taxonómica¹³ de los organismos en estudio a partir del espesor de la Banda, que se quiere a la vez confirmar la hipótesis, referente al espesor de la Banda, ya que se sostiene que el ancho de la Banda tiene directa relación proporcional con el espesor de la muestra de esa secuencia en particular, ya que cada Banda corresponde a una secuencia. Además se podrá concluir con ello el número de especies que distintas que existen en un hábitat en particular

12 Aplicaciones utilizadas para el mejoramiento de la calidad de imagen.

13 Ver Glosario.

y hacer la comparación con los demás, a través de las Bandas detectadas entre los diferentes carriles.

En resumen, los principales problemas detectados son los relacionados con la Detección e Identificación de Bandas y su Cuantificación, temas por los cuales se abordarán en las siguientes secciones del documento.

3.2.1. Descripción del Problema de Detección e Identificación de Bandas.

Como se mencionó en el análisis de la situación actual, uno de los problemas del laboratorio consistía en la falta de precisión y de sensibilidad a la hora de detectar y cuantificar las bandas del DGGE, por parte del programa Quantity -One, ya que este opera a través de estándares, los cuales conllevan a arrojar resultados con algún porcentaje de falló, estos errores se pueden observar en las imágenes que se visualizan en la Figura 11. Se puede ver que en los carriles la detección automática es deficiente, ya que el usuario deberá de borrar lo que su aparecer no corresponde a Banda, lo que en este caso corresponderían las Bandas con los siguientes números: 8-9-12-13-14-15, lo que daría en este carril una probabilidad de fallo del 38%, en el mejor de los casos por carril, el cual se quiere disminuir con el módulo DIBD

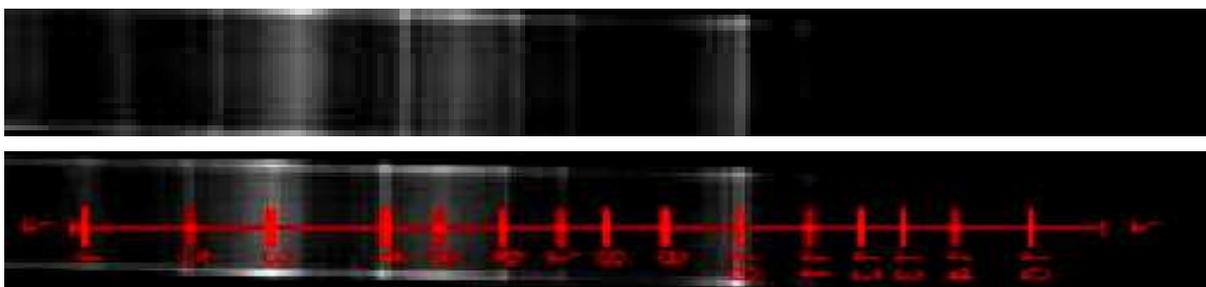


Figura 11: Arriba Carril original. Abajo: Bandas detectadas por el Q1

En la Figura 12 y 13 se puede apreciar de manera completa lo que se mencionó como deficiencia en el procesamiento de una imagen DGGE, en donde se aprecia las opciones para que el usuario pueda eliminar o agregar las bandas que están demás o las que no fueron posibles detectar, a través del mouse.

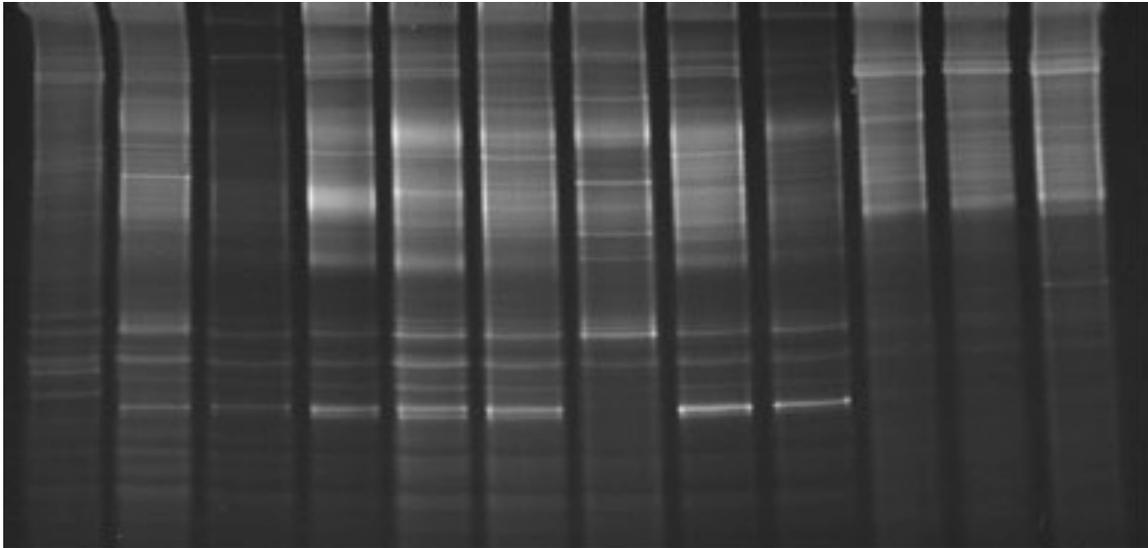


Figura 12: Imagen DGGE1_1

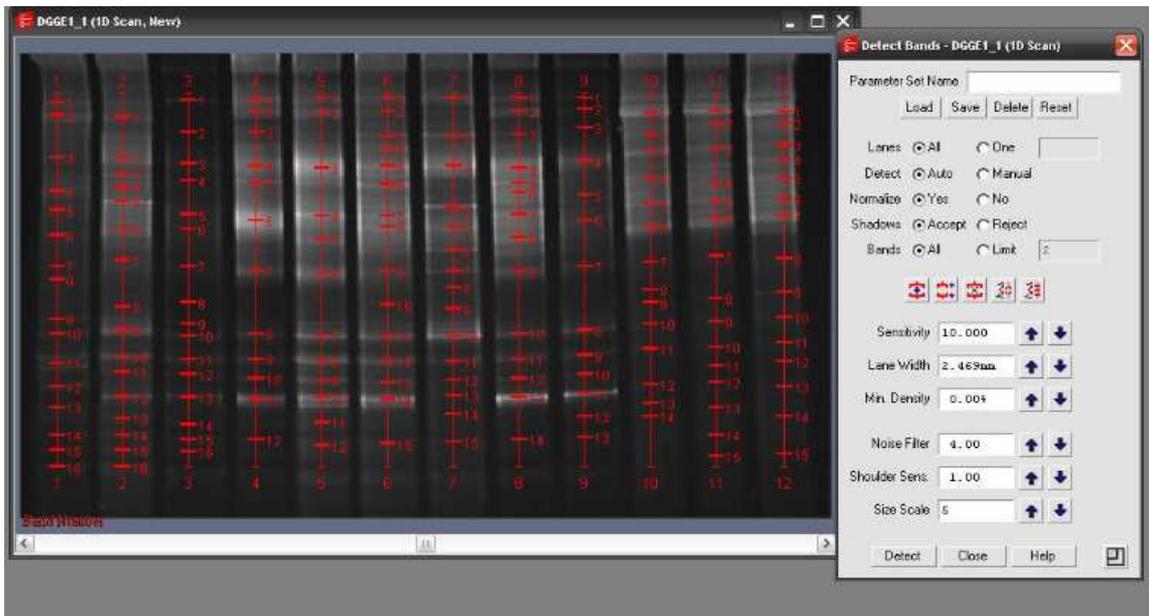


Figura 13: Procesamiento en la Aplicación Quantity One de la imagen de la Figura 15

Otro problema que se pudo observar con respecto a la detección de Bandas por parte del Q1, es el hecho de que en vez de detectar las Bandas, éste detecta el background o el fondo

de la imagen como se observa en las imágenes de la Figura 14, situación que se repitió cuando se procesó la imagen que sirvió de prueba en la etapa de proyecto de título.

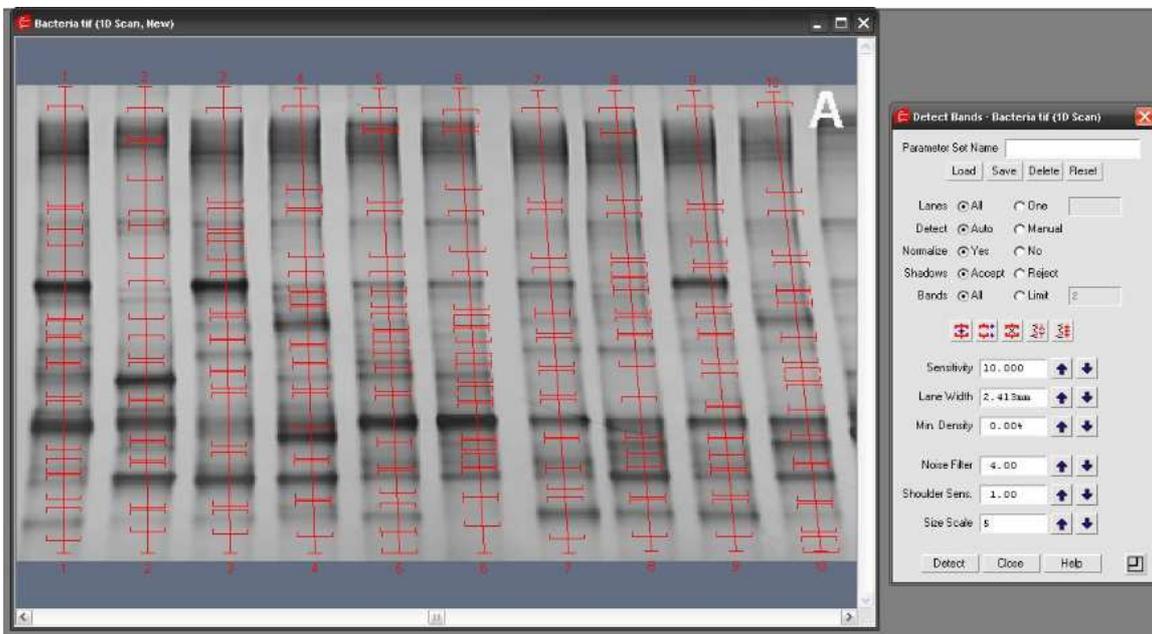
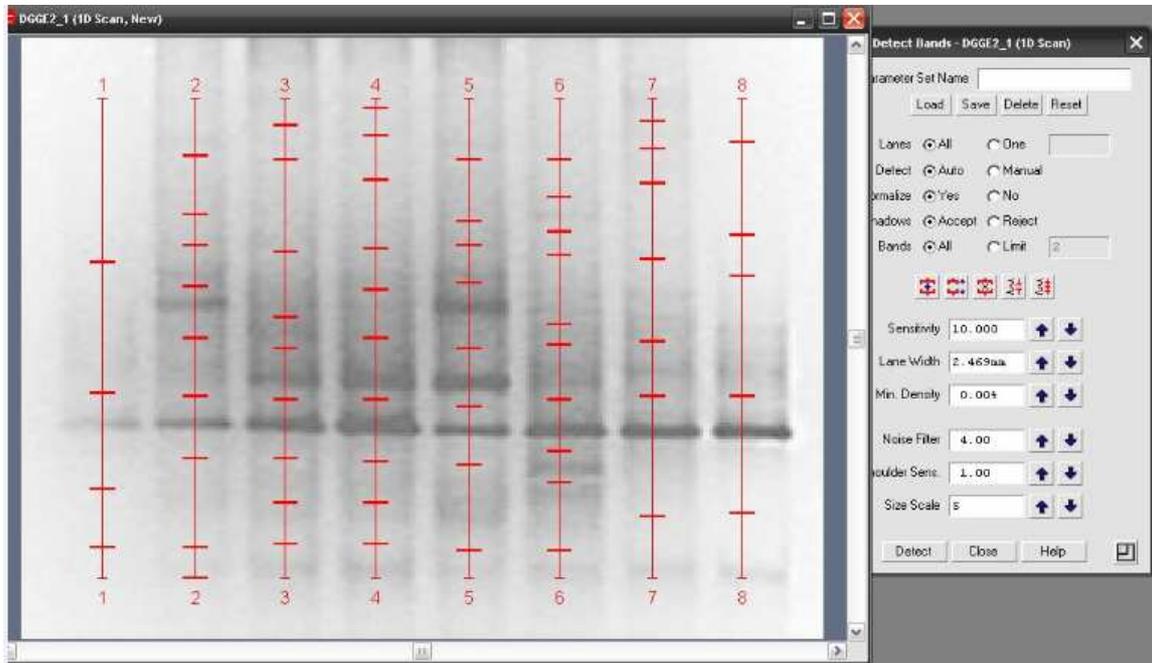


Figura 14: Problemas en la detección de Bandas con la Aplicación Quantity One

3.2.2. Descripción del Problema de Cuantificación de Bandas.

La cuantificación de Bandas, más que un problema visto por el laboratorio es una oportunidad, debido a que hasta el momento el mundo de la Biología, tiene dado por certeza, el hecho de que la técnica del DGGE es de carácter cualitativo, es decir, no existe hasta el día de hoy una manera en cómo medir a través de cifras matemáticas las secuencias de hebras de ADN que componen cada banda, todas ellas en base al peso molecular. Debido a que se quiere comprobar además la hipótesis, que hace referencia al ancho de cada Banda, si existe relación con el espesor de la muestra recogida originalmente [7].

Se puede agregar como un segundo factor el hecho de que Quantity-One, no contiene entre sus aplicaciones esta funcionalidad, lo que fundamentaría lo señalado anteriormente, ya que sus principales funciones en base a las bandas se resumen de la siguiente manera y representado visualmente en la Figura 15.

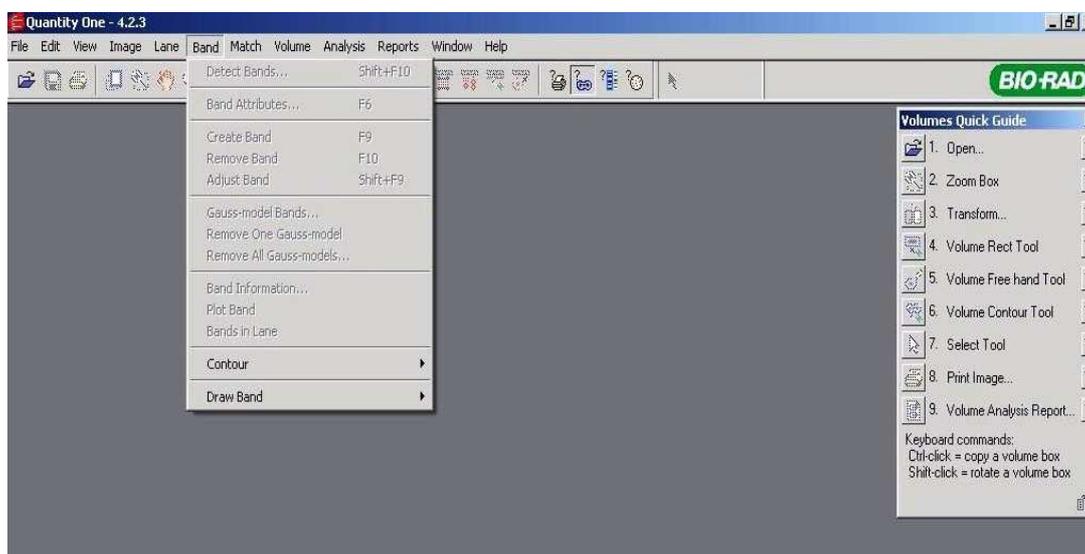


Figura 15: Aplicaciones referente a las Bandas

3.3. Definición de requerimientos.

Los requerimientos son la pieza fundamental, en cualquier tipo de proyecto de software, sobre todo en áreas como la Biología, en donde el informático tiene mucho que aprender para poder satisfacer sus necesidades, incluso si éste fuera experto en la Bioinformática. Esta labor debe ser efectuada en conjunto con quienes presencian el problema a diario, señalado en la sección 3.2. de este documento. Logrando con ello un sistema más robusto y de mayor calidad. Sin embargo, para conseguirlo se debe de tomar en cuenta lo siguiente:

- Obtener la información que los usuarios desean o lo que esperan del sistema.
- Identificación de los actores.
- Especificación de las distintas clases de requerimientos.

3.3.1. Restricciones.

Esta sección permite al lector identificar las principales restricciones del sistema, en base al diseño y a su construcción, como por ejemplos el lenguaje de programación, las herramientas Case, requerimientos del software, por mencionar algunos.

En este sistema, las restricciones encontradas son mínimas, las cuales se pueden resumir de la siguiente manera:

- El sistema deberá basarse en las normas y estándares de calidad de software como: Estándares ISO/IEC 9126-1[8]
- El entorno de trabajo o las tecnologías de implementación deberán ser las apropiadas para el desarrollo del sistema CYDB.
- Deberá integrarse de manera correcta a los demás módulos en construcción y al Sistema General DGGE.

Con respecto a la administración de la información proporcionada por el sistema, será interpretada por quien haga uso de ello, ya que sólo se encargará de procesar la imagen,

pero los resultados arrojados serán interpretados por el usuario o por otro módulo, como el de SDC.

3.3.2. Identificación de actores

El sistema de Detección, Identificación y Cuantificación de Bandas, considera solamente a:

- **Usuario Final**

El usuario final está representado por todos los investigadores del área de la Biología, Bioquímica o Farmacéutica, que concurren al laboratorio, sin importar que éstos sean académicos o alumnos, ya que no existe jerarquía en los roles, ni muchos menos privilegios en el manejo del sistema.



Figura 16: Usuarios Finales



- **Otros Módulos**

Otro actor identificado son los demás módulos, como por ejemplo: el Sistema de Interfaz o SDI , que hara uso de la información que proporcionará este sistema, dejando en claro además que éste al igual que los demás conformaran el Sistema General DGGE.

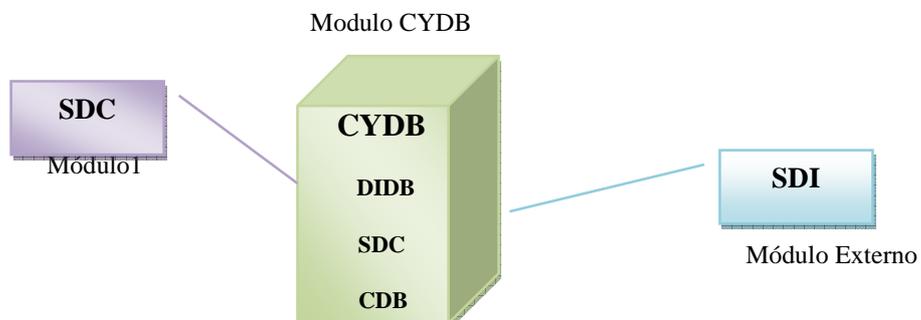


Figura 17: Comunicación con otros Módulos

3.4. Especificación de requerimientos

La especificación de requerimientos, aborda los temas relacionados con los requisitos que debe de cumplir el sistema, para poder satisfacer las necesidades de los usuarios del laboratorio, y así poder diseñar como se dijo en un principio un sistema de mejor calidad.

3.4.1. Requisitos de interfaces externas

La interfaz se define como el lugar en donde el sistema interactúa con los usuarios y con los demás módulos. En la cual se especifica el tipo de información que recibe y que envía al exterior.

- *Interfaces de Usuario*

Para el sistema a construir no está establecido crear interfaces gráficas, por lo que este requisito no es un determinante, para su implementación. Sin embargo, se requiere determinar la manera en cómo el sistema deberá generar los reportes de los sub- módulos.

- *Interfaces de Hardware*

El sistema puede funcionar en correctamente en plataformas Windows.

- *Interfaces de Software*

Para la ejecución del sistema, aparte de la instalación del sistema deberán de agregarse al computador las librerías pertinentes a la herramienta de trabajo MATLAB para el buen funcionamiento de la aplicación

3.4.2. Requisitos Funcionales.

Los requisitos funcionales permiten describir lo que el sistema debe de hacer, con respecto a las funcionalidades del software, las cuales son obligatorias para la etapa de implementación.

Funciones Básicas del Sistema.

Referencia	Descripción del Requerimiento
R01	El sistema debe permitir al usuario, realizar la Detección e Identificación de las Bandas de una manera eficiente.
R02	El sistema debe permitir al usuario, realizar la Cuantificación de Bandas, de una manera eficiente.
R03	El sistema deberá permitir su uso por cualquier usuario del laboratorio, sin necesidad de la utilización de contraseñas o claves de accesos.
R04	El sistema deberá generar reportes, mostrando en ellos los resultados correspondientes a la cuantificación, para que los usuarios tomen nota de ello.
R05	Permitir la captura de cualquier tipo de imagen digital en cualquier formato.
R06	Generar un reporte de las Bandas detectadas e identificadas y cuantificadas

Atributos Generales

Atributo	Detalle y Restricción de Frontera
Restricciones de Tiempo	Cada aplicación del sistema como la Identificación de Bandas, debe ser ejecutada en un rango menor que lo esperado por el Quantity One, el cual tarda alrededor de 30 segundos aproximadamente.
Restricciones de Acceso	Todos los usuarios pueden utilizar el programa para fines académicos o para sus investigaciones.
Sistema Operativo	No hay restricciones para especificar el sistema operativo.
Interfaz Gráfica	Se hará uso de ello, en el momento que se generen los reportes concernientes a la detección e identificación, los cuales serán vistos por pantalla, por ejemplo a través de un archivo de texto.

Atributos relacionados con cada Función.

Ref	Función	Atributo	Detalles y Restricciones
R01	El sistema debe permitir al usuario, realizar la Detección e Identificación de las Bandas de una manera eficiente	Restricciones de Tiempo	Debe ser menor a 30 segundos
		Restricciones de accesos	No se permiten
R02	El sistema debe permitir al usuario, realizar la Cuantificación de las Bandas de una manera eficiente.	Restricciones de Tiempo.	Debe ejecutarse en un tiempo menor a 30 segundos.
		Restricciones de accesos	No se visualizan

R03	El sistema deberá permitir su uso por cualquier usuario del laboratorio, sin necesidad de la utilización de contraseñas o claves de accesos	Restricciones de acceso	No se permiten
R04	El sistema deberá generar reportes, mostrando en ellos los resultados correspondientes a la cuantificación, para que los usuarios tomen nota de ello.	Interfaz Gráfica	Debe ser capaz de entregar la información pertinente a la cuantificación de Bandas, dentro de un archivo por ejemplo en uno con extensión texto.
R05	Permitir la captura de cualquier tipo de imagen digital en cualquier formato.	Restricciones de tiempo	Debe de ser en un tiempo menor a 5 segundos
R06	Generar un reporte de las Bandas detectadas e identificadas.	Interfaz Gráfica	Debe ser capaz de entregar la información pertinente a la Detección de Bandas, dentro de un archivo por ejemplo en uno con extensión txt.

3.4.3. Requisitos No Funcionales.

Los Requisitos No Funcionales, describen las restricciones del sistema con respecto a:

Rendimiento

El sistema deberá de ser lo bastante rápido y eficiente, en la realización de sus operaciones, ya que esto permitirá establecer una ventaja considerable con respecto al software utilizado por el laboratorio hasta el momento, todo ello en base a la manera automática en que ambos operan.

Seguridad

La aplicación a realizar no necesita de mayor seguridad que la entregada por el propio establecimiento, debido a que no maneja datos o información confidencial.

Mantenimiento

El sistema será desarrollado con el fin de que se pueda lograr una buena mantención además de corregir y perfeccionar el sistema. Revisiones en forma mensual por un período de 6 meses, para verificar el buen funcionamiento del sistema.

Respaldo

La información contenida en el sistema deberá guardarse en dispositivos de almacenamientos secundarios como: pendrive o CD.

Usabilidad

El tiempo que tomará instruir a los usuarios en la adaptación al nuevo sistema será de una semana. Adjunto al programa, irá un manual de usuario que contendrá todas las explicaciones de cómo usar el sistema y con las respectivas preguntas frecuentes que se podría plantear.

3.5. Componentes del Sistema.

Este diagrama representa al módulo CYDB, con los principales componentes para su implementación, los cuales se encuentran en el rectángulo del centro de la Figura 18.

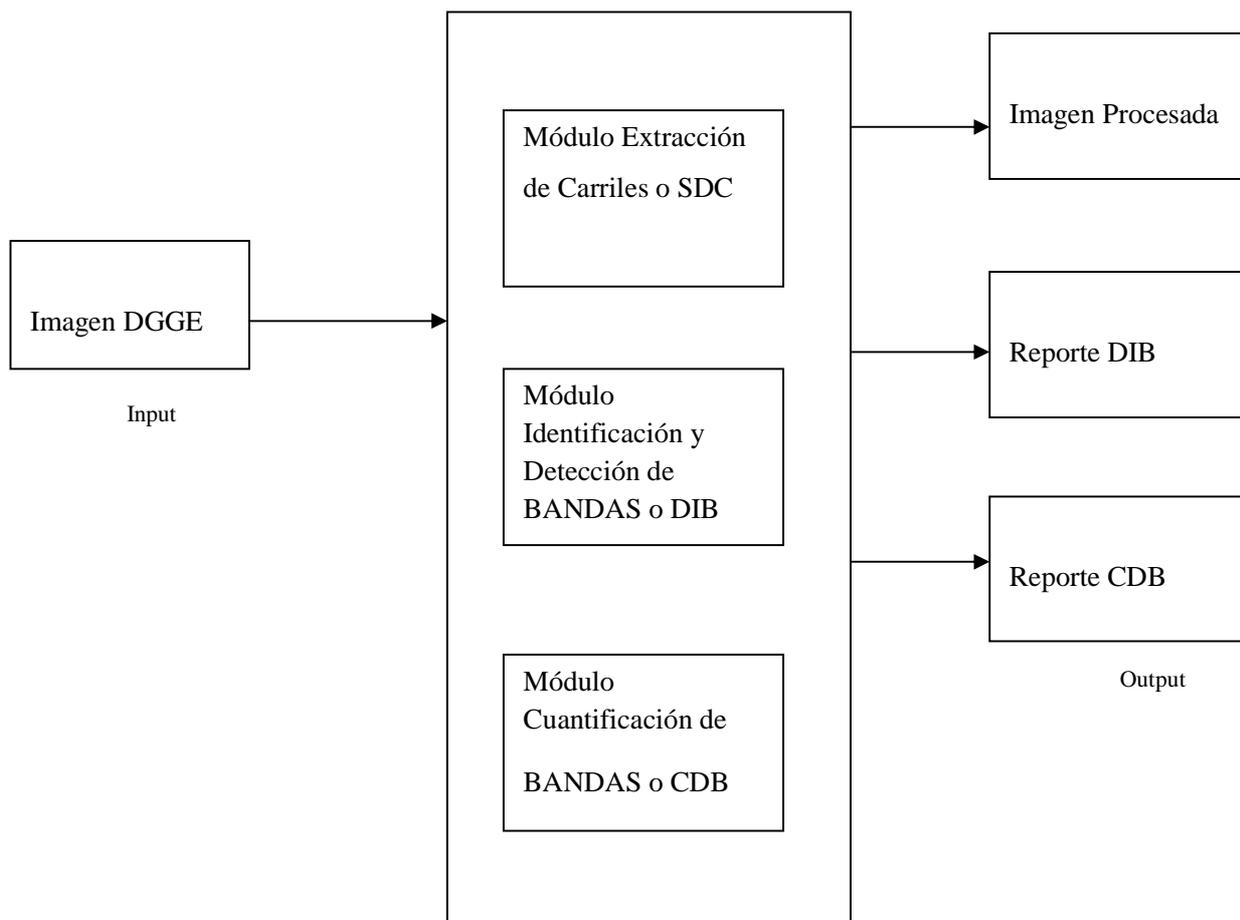


Figura 18: Esquema del Funcionamiento del Módulo CYDB.

3.6. Diagrama de Paquetes y Arquitectura del Sistema

3.6.1 Diagrama de Paquetes

El diagrama de paquetes permite organizar de manera lógica el sistema CYDB, el cual puede ser demostrado en la Figura 19 y detallado en la Figura 20.

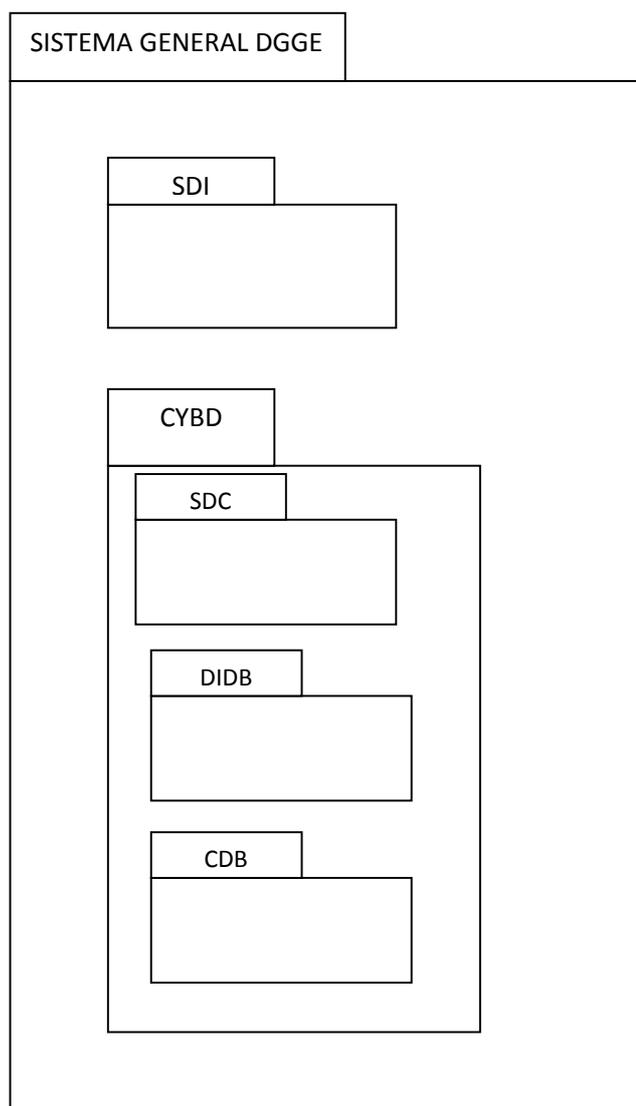


Figura 19: Diagrama de Paquetes

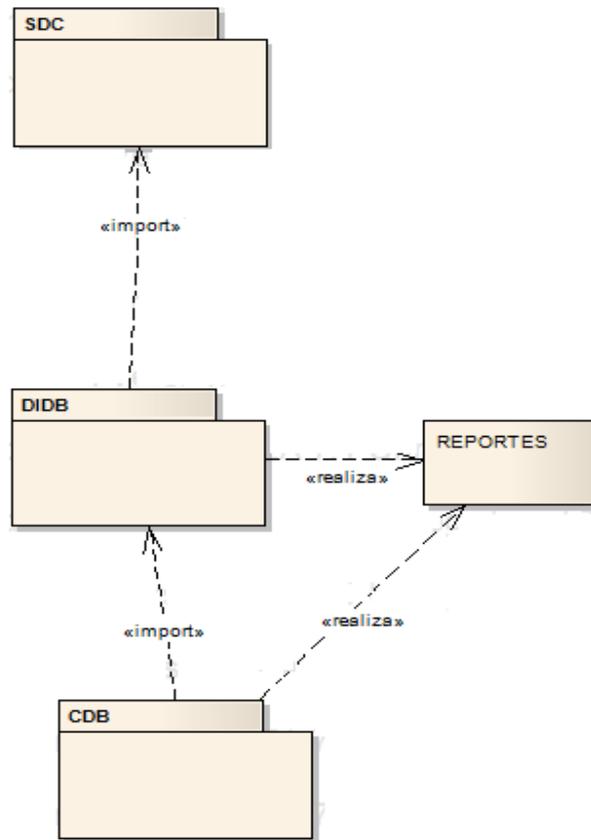


Figura 20: Diagrama de Paquetes en detalle.

3.6.2. Arquitectura del Sistema

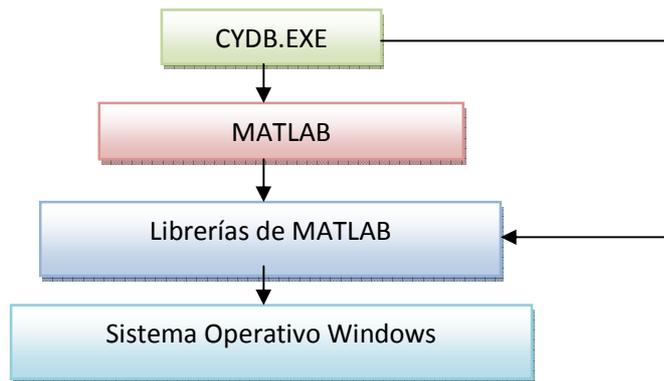


Figura 21: Arquitectura del Sistema

3.7. Diagrama y Casos de Uso

Diagrama: Extracción de Carriles

Este diagrama a pesar de no representar a algún requerimiento preestablecido, se ha diseñado debido a la falta de algún modulo externo que permitiera realizar la extracción de los carriles para facilitar con ello el proceso de Identificación y Detección de Bandas.

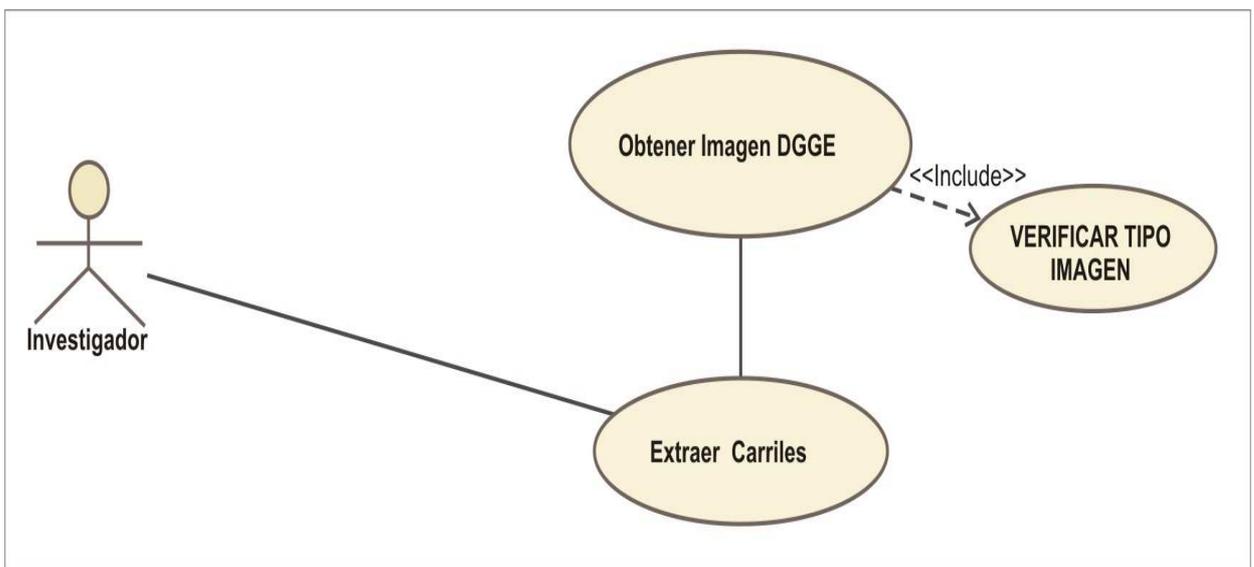


Figura 22: Caso de Uso del Sub-Módulo DIDB

Nombre	Módulo de Segmentación de Carriles
Descripción	Permite la extracción de carriles a partir de la obtención de cualquier tipo o formato de imagen.
Actores	Módulo de Detección e Identificación de Bandas o DIDB.
Precondiciones	La imagen a procesar debe estar guardada en formato digital en algún lugar del equipo.
Flujo Normal	<ol style="list-style-type: none"> 1. El módulo DIDB inicia su ejecución. 2. El módulo captura la imagen DGGE, la cual puede estar almacenada en cualquier tipo o formato digital: JPEG, TIF, RGB, Escala de Grises. 3. Los carriles son detectados y extraídos. 4. Una vez extraídos son almacenados en un nuevo archivo, y listos para ser usados por el módulo DIDB.
Postcondiciones	Los carriles deben de estar segmentados, para que el módulo DIDB los pueda procesar cada uno de ellos de manera individual.

Diagrama: Identificar y Detectar Bandas.

Este diagrama representa a los R01 y R06

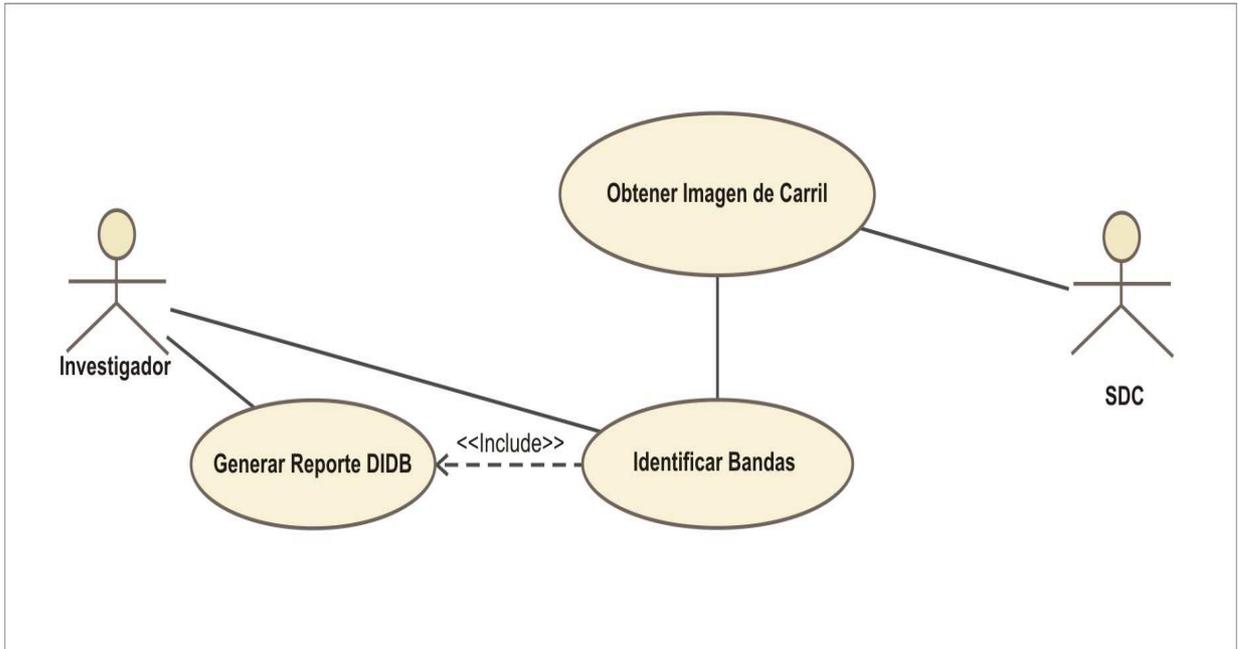


Figura 23: Caso de Uso del Sub-Módulo DIB

Caso de Uso: Identificar y Detectar Bandas

Nombre	Identificar y Detectar Bandas
Descripción	Permite identificar y detectar bandas, a partir de una imagen en cualquier formato, como por ejemplo: jpeg, jpg, gif,etc. La cual corresponde a un carril en particular.
Actores	Investigador, Modulo de Segmentación de Carriles.
Precondiciones	Antes de comenzar a ejecutarse el carril debe estar representado de manera individual.
Flujo Normal	<ol style="list-style-type: none"> 1. El investigador inicia su ejecución. 2. El módulo captura cada carril por separado, a través de otro modulo que permite ese procedimiento, que se explicará detalladamente en el capítulo IV del documento. 3. El carril es analizado, en donde se le aplican procedimientos para disminuir el ruido. 4. Una vez mejorado se identifican las posiciones de las Bandas, las cuales. 5. El sistema genera la imagen del carril con las Bandas detectadas y las posiciones, todo ello en un reporte.
Postcondiciones	El investigador o usuario final, debe de recibir un reporte con los resultados obtenidos. Para sus posteriores interpretaciones y compararlos con el Quantitty-One.

Diagrama Caso de Uso: Cuantificar Bandas.

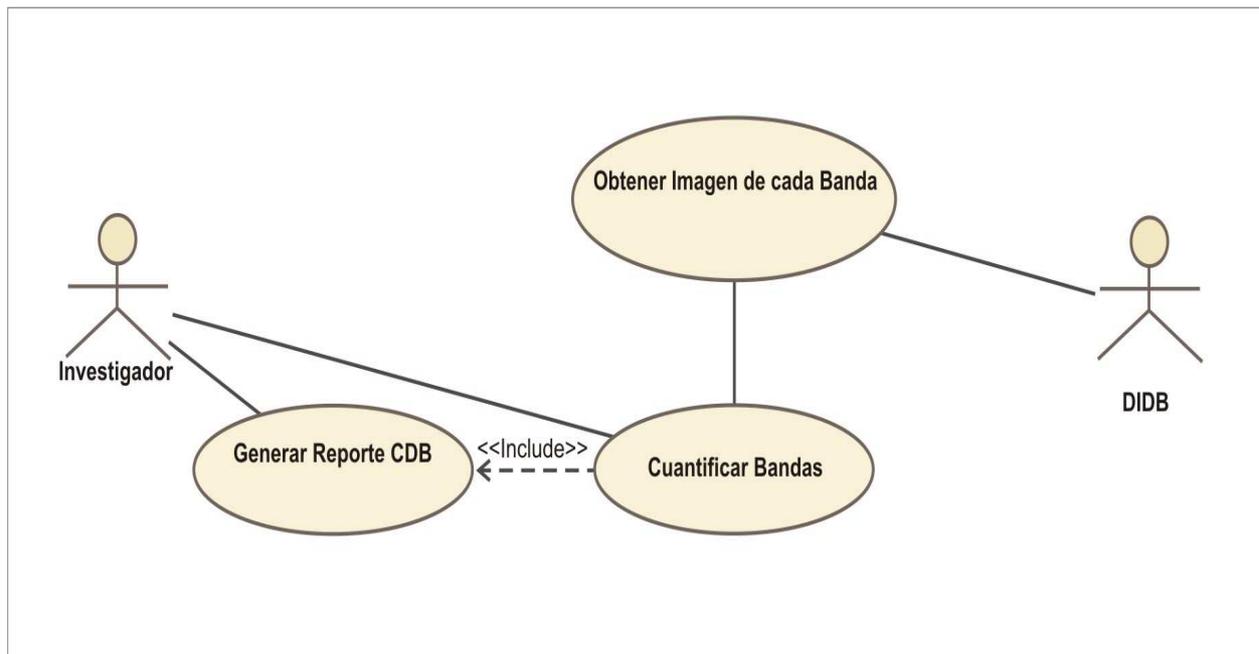


Figura 24: Diagrama Caso de Uso Sub-Módulo CDB.

Caso de Uso: Cuantificar Bandas

Nombre	Cuantificar Bandas
Descripción	Permite cuantificar las bandas detectadas, a partir de una imagen en cualquier formato, como por ejemplo: jpg, gif, etc. La cual corresponde a un carril en particular.
Actores	Investigador, Modulo de Detección e Identificación de Bandas.
Precondiciones	Antes de comenzar a ejecutarse las posiciones de las bandas deben de estar identificadas.
Flujo Normal	<ol style="list-style-type: none"> 1. El investigador inicia su ejecución. 2. El módulo lee cada carril por separado, con las posiciones de las bandas identificadas, a través del modulo de Identificación y Detección de Bandas. 3. Una vez identificada las posiciones, se buscan los límites de cada banda. 4. Cada banda es separada del carril, las cuales se vuelven a analizar para disminuir el ruido. 5. Ya separadas, el modulo procede a cuantificarlas a partir de la intensidad de sus pixeles, en base de estimadores estadísticos. 6. El sistema genera un reporte el cual es visto por el investigador, en el cual se resume el valor o cuantificación de cada Banda, para su posterior interpretación.
Postcondiciones	El investigador o usuario final, debe de recibir un reporte con los resultados obtenidos. Para sus posteriores interpretaciones.

3.8. Análisis de riesgo.

Número	1
Nombre	La imagen DGGE ingresada no sea de calidad.
Tipos de Usuario	Investigador o Usuario Final
Probabilidad de Ocurrencia	0.01 en el momento de ejecución.
Justificación	Si la imagen es difícil de leer y analizar producto del descuido o de ruido imposible de disolver, el módulo no podrá operar eficientemente.
Medidas de Solución	Que exista por parte de los usuarios del sistema una mayor preocupación en el desarrollo de la Imagen antes de analizarla.

Capítulo IV:

Trabajo desarrollado.

En el siguiente capítulo se especificará paso a paso el desarrollo del Sistema o Módulo CYDB, su implementación, la manera de generar reportes y las pruebas a distintas imágenes DGGE, además de comparar los resultados obtenidos con los del Quantity-One.

4.1. Definición de las Actividades del Sistema CYDB

El Sistema CYDB o de Cuantificación y Detección de Bandas entre sus principales actividades se encuentran resumidas de la siguiente manera y representadas en el diagrama de flujo señalado en la Figura 25.

1. Capturar imagen DGGE de cualquier tipo o formato.
2. Extracción de Carriles.
3. Identificación de Bandas.
4. Cuantificación de las Bandas.
5. Generación de Reportes por cada imagen DGGE en formato html.

En donde cada una de ellas será detallada, en las siguientes secciones del presente capítulo.

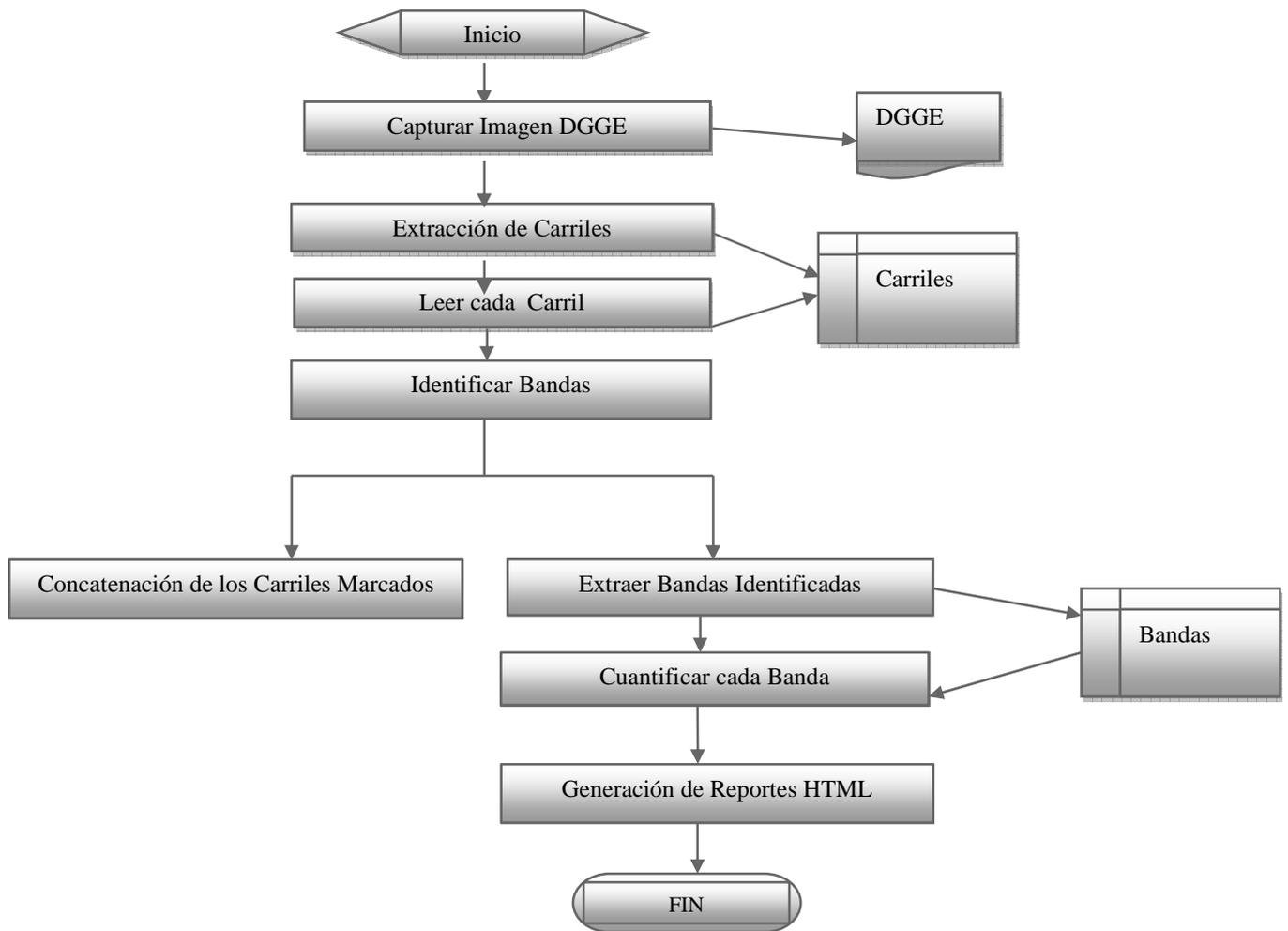


Figura 25: Diagrama de Flujo de las actividades del Sistema general

4.2. Definición General de los Sub-Módulos de CYDB

Como se pudo observar a grandes rasgos en los capítulos anteriores, el Sistema CYDB para su construcción se tuvo que dividir en dos Sub-Módulos los cuales, a la vez requieren de funciones específicas para su desarrollo y ejecución. Los cuales se nombran de la siguiente manera:

1. *Sub-Módulo 1: “Detección e Identificación de Bandas o DIDB”*

- Función 1: Obtener Negativo → `funcion_negativoRGB.m`
- Función 2: Extraer Carriles → `función_extraecarriles.m`
 - Sub-función: Detecta posiciones de los carriles → `find_gelpos.m`
- Función 3: Normalización de Histograma → `normaliza_histograma.m`

2. *Sub-Módulo 2: “Cuantificación de Bandas o CDB”*

- Función 1: Extracción de Bandas → `función_extraebandas.m`
 - Sub-Función: Detectar posiciones bandas → `función_bordes_banda.m`
 - Sub-Función: Extrae las bandas → `función_extrae_bandas.m`
- Función 2: Normalización Histogramas Bandas → `Normaliza_histBand.m`

Las cuales son llamadas en un programa principal reconocido como `programa_principal.m`, la cual además aparte de las funciones que mencionadas anteriormente, se encuentra la que originaría la generación de reportes.

4.3. Desarrollo del Sub-Módulo DIDB

Para el desarrollo del Sub-Módulo de Detección e Identificación de Bandas en Imágenes DGGE, se debe de tomar en cuenta los siguientes pasos, los cuales se resumen en el siguiente diagrama de flujo representado en la Figura 26.

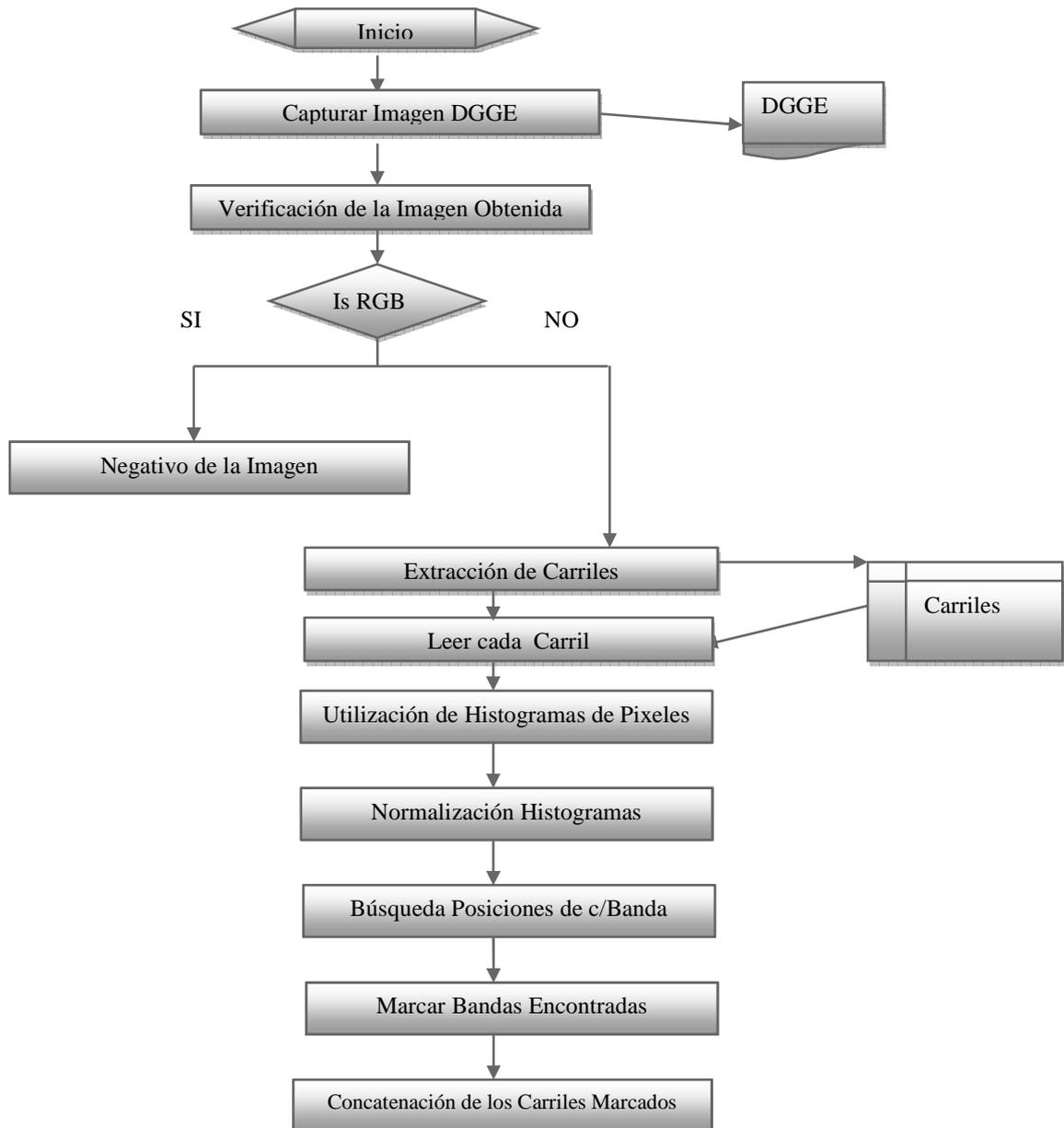


Figura 26: Diagrama de Flujo de los Pasos del Sub-Módulo DIDB

Paso 1: Captura de la Imagen:

Antes de ejecutar este sub - módulo el primer paso que se debe realizar es la obtención de la imagen DGGE desde archivo o algún lugar del equipo, la cual puede estar en cualquier formato digital, es decir, JPEG, TIF o PNG, o de cualquier tipo: en escala de grises o a color RGB, debido a la creación de una función, que permite cambiar el formato y obtener el negativo de la imagen a color. En la Figura 27, se puede observar una imagen DGGE a color, la cual ha sido capturada por este sub-módulo con la intención de demostrar en los siguientes pasos, la obtención de su negativo el cual servirá de base para la extracción de carriles y detección de las bandas, por el hecho de trabajar con los umbrales de luminiscencia entregados por los valores cercano a la tonalidad del blanco.

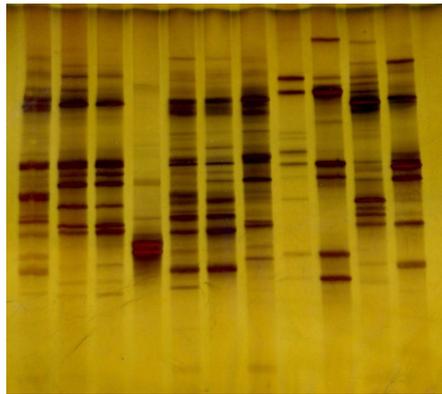


Figura 27: Imagen DGGE a Color

Desde la perspectiva del código, la manera de obtener la imagen desde archivo sería de la siguiente manera

Alternativa 1

```
fileName = 'IMAGENES_DGGE/DGGE1';  
imageName = strcat(fileName, '.jpeg');  
A = imread(imageName);
```

Alternativa 2

```
fileName = 'IMAGENES_DGGE/DGGE1';
imageName = strcat(fileName, '.tif');
A = imread(imageName);
```

En donde:

fileName: Corresponde a la ruta del directorio en donde está almacenada la imagen DGGE digitalizada.

imageName: Realiza la conversión con la ayuda de la función *strcat* (función Matlab) la concatenación de la ruta con el formato en el que ha sido guardado, el cual puede ser *TIF* o *JPEG*, como se observa en la alternativa 1 y 2.

A: Es la variables tipo matriz que almacena la imagen a través del comando *imread*.

Paso 2: Verificación de la Imagen Obtenida

Una vez que la imagen haya sido obtenida y almacenada en la variable *A*, se verifica el tipo de imagen, comprobando si corresponde al tipo RGB o al de Escala de Grises, ya que al ser de este último tipo no se realizaría ninguna transformación, es decir, no se obtendría el negativo de *A*, porque todas las de este tipo vienen con estas características al ser obtenidas mediante una cámara de luz ultra violeta, un ejemplo de ello es el que se aprecia en la Figura 28.

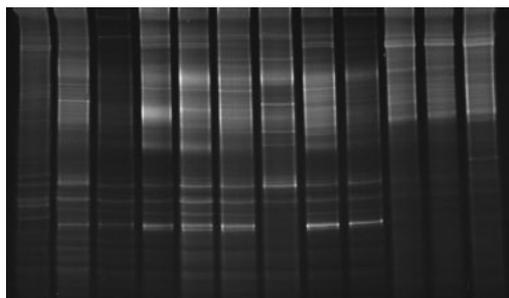


Figura 28: Imagen capturada a través de una cámara UV.

En cambio, si la imagen hubiera sido del tipo RGB, se debe obtener su negativo a través de la siguiente función denominada como: *funcion_negativaRGB.m*.

La cual opera de la siguiente manera, desde la perspectiva de la programación:

```

if (isrgb(A) == 1)
    C = funcion_negativoRGB(A);
else
.....
End
    
```

En donde:

ISRGB: Función que permite verificar si la imagen obtenida corresponde a una del tipo RGB, si es verdad será igual a 1.

C: Variable que almacena la imagen que ha sido modificada por *funcion_negativoRGB.m*, la cual se encarga obtener el negativo de *A*, para facilitar la tarea de la identificación de las bandas y extracción de carriles, ya que para ello se ha hecho uso de los umbrales de luminiscencia entregados por las tonalidades cercanas al blanco.

Cabe señalar que todo el procedimiento descrito anteriormente, puede ser representado en la Figura 29:

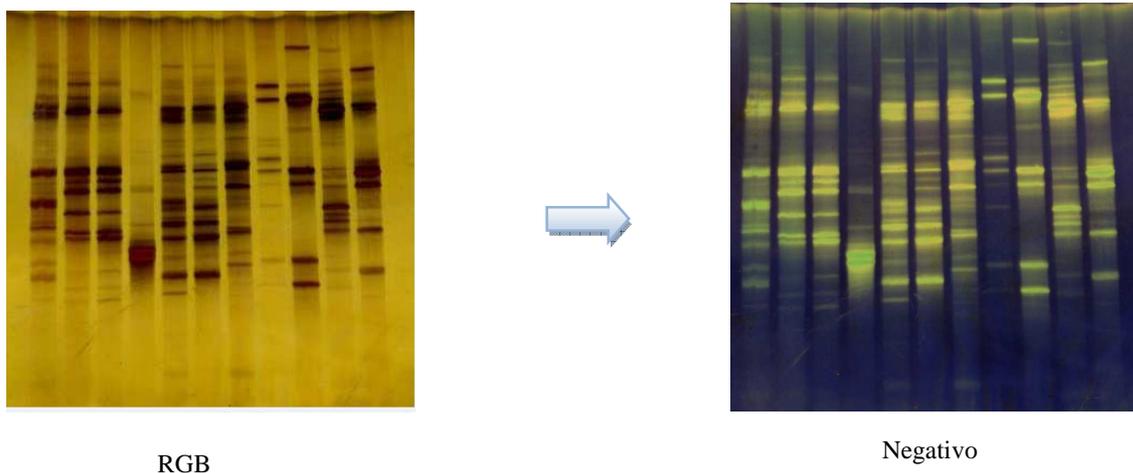


Figura 29: Obtención del Negativo de una Imagen RGB.

Paso 3: Extracción de los Carriles

Una vez que la imagen se encuentre en las condiciones dadas por la sección anterior, se puede dar paso a la extracción de los carriles, a través de *función_extrae_carriles.m*, la cual identifica en primer lugar las posiciones de los límites de cada carril para poder extraerlos y guardarlos en un nuevo directorio, el que sería una sub-ruta dentro de la misma ruta de la imagen original.

La manera en cómo opera esta función se describe a continuación:

```
function [gel] = funcion_extrae_carriles(C,ncarril,nrefe)
    ngel = ncarril;
    nref = nrefe;
    B = C(:, :, 1);
    [nx,ny] = size(B);
    gelpos = findGelPos(B,ngel);
    fileName = 'MAS_DGGE/DGGE5/';

%% CARRIL 1
C1 =B(:,gelpos(1,1): gelpos(1,2));
imwrite(C1,strcat(fileName,'1.jpg'),'JPEG');
figure(3)
imshow(C1)

%% todo este ciclo se repite hasta el último carril, que en este caso es
el carril 11
```

En donde:

[gel] = funcion_extrae_carriles(A,ncarril,nrefe): Recibe los parámetros desde el programa principal (main), en donde *C* es la imagen que ha sido modificada en el paso anterior, *ncarril* es la variable que contiene el número de carriles y *nrefe* almacena el número de carriles de referencia.

[nx,ny] = size(B): El vector *[nx,ny]* se encarga de almacenar el tamaño correspondiente al ancho y al alto de la imagen *B*, a través del comando *size*.

gelpos = findGelPos(B,ngel): La variable *gelpos* es una matriz que contendrá las posiciones de los límites inferiores y superiores de los carriles, a través de la función *findGelPos*, el cual recibe como parámetros a la imagen guardada en la variable *B* y los *ngel* o el número de carriles.

$C1 = B(:, gelpos(1,1): gelpos(1,2))$: La variable $C1$ corresponde al primer carril de la imagen B , en donde es segmentado en base al siguiente orden de la matriz $B(:, gelpos(1,1): gelpos(1,2))$, ya que el primer valor antes del punto y coma se refiere al ancho del carril, $gelpos(1,1)$ y $gelpos(1,2)$ límites inferiores y superiores del carril 1.

$imwrite(C1, strcat(fileName, '1.jpg'), 'JPEG')$: El comando $imwrite$ se encarga de guardar en el archivo $filename$ el carril 1 ($C1$) con la extensión jpeg, a través del comando $strcat$ que transforma $C1$ a ese tipo de imagen.

Todo lo expresado en este paso se puede visualizar de manera gráfica en la Figura 30, 31 y 32

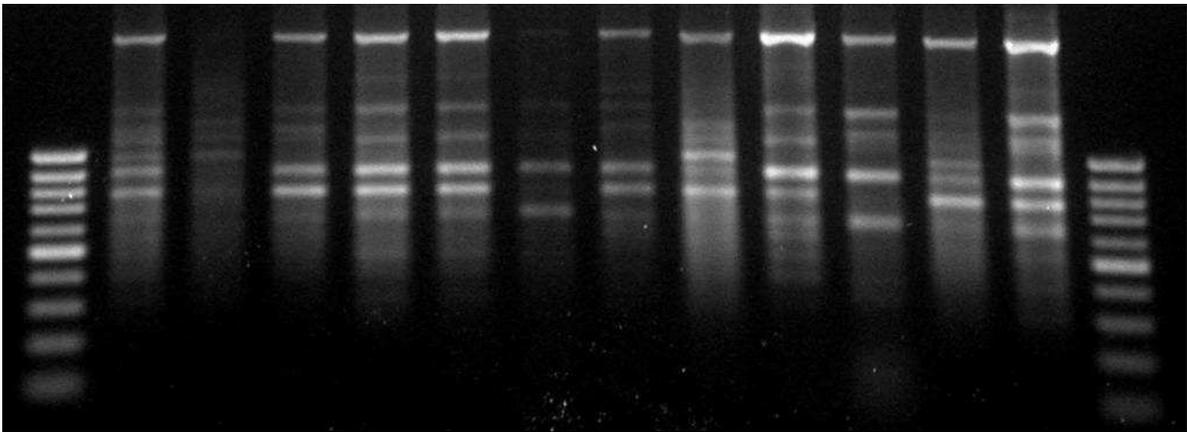


Figura 30: Imagen obtenida desde el paso 2.

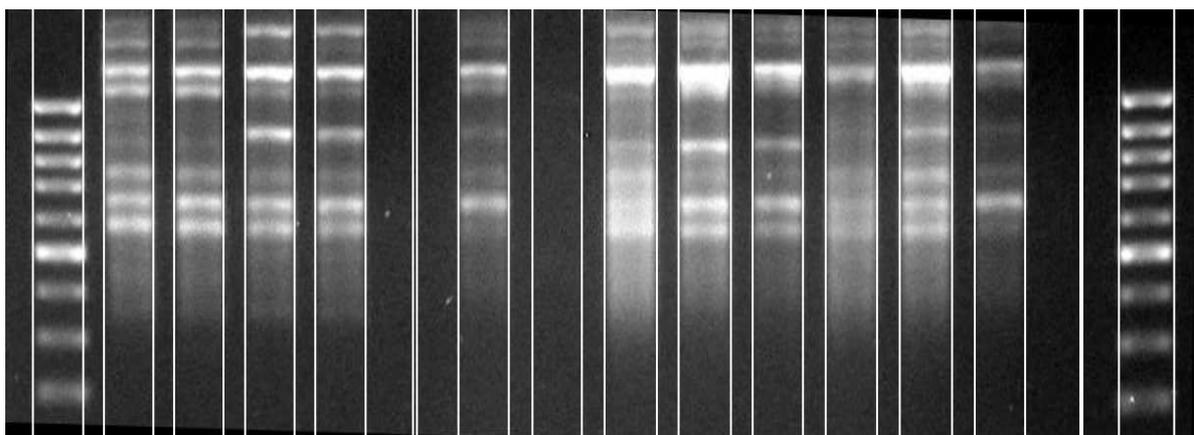


Figura 31: Identificación de los límites de cada carril (gelpos)

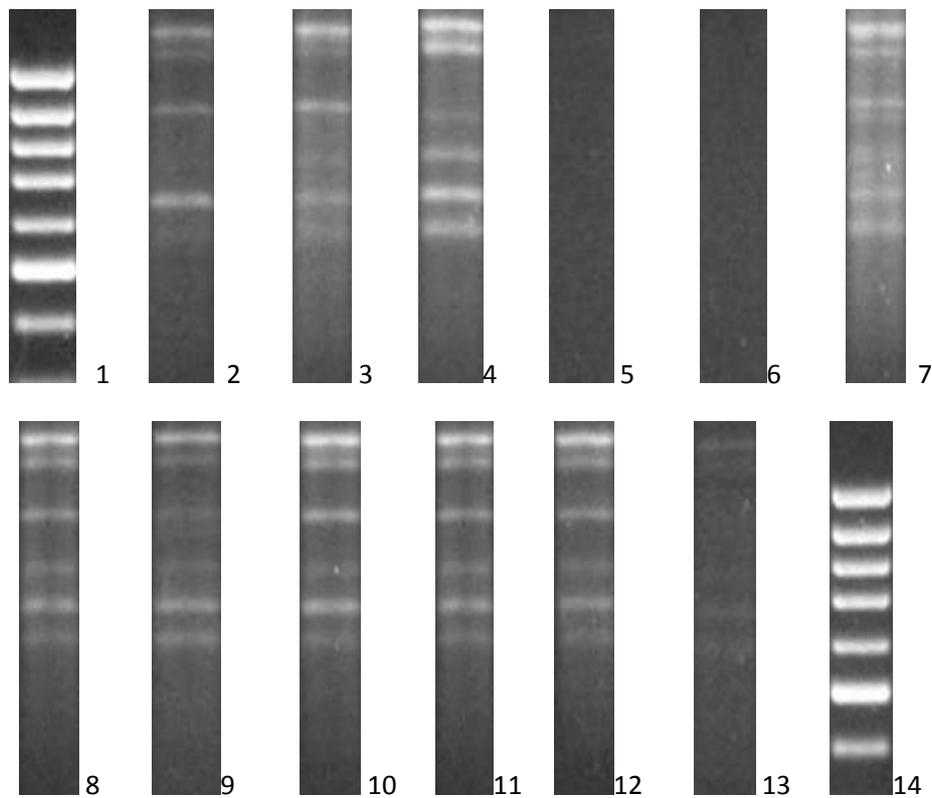


Figura 32: Extracción de los carriles.

Paso 4: Utilización del Histograma de Suma de Píxeles y Suavizamiento

Una vez que se hayan extraído todos los carriles de las imágenes DGGE, es posible comenzar con la detección de las bandas. La cual es inicializada desde la lectura de los parámetros a partir del programa principal (main.m), como se muestra a continuación:

```
C1 = funcion_didb(A1) %se invoca la función función_didb.m
```

```
function gel = funcion_didb(A) %recibe la imagen correspondiente al
primer carril como parámetro devolviendo al carril con las bandas
identificadas.
```

Al momento de obtener el carril, es girado horizontalmente quedando de la siguiente manera como se expresa en la Figura 33.

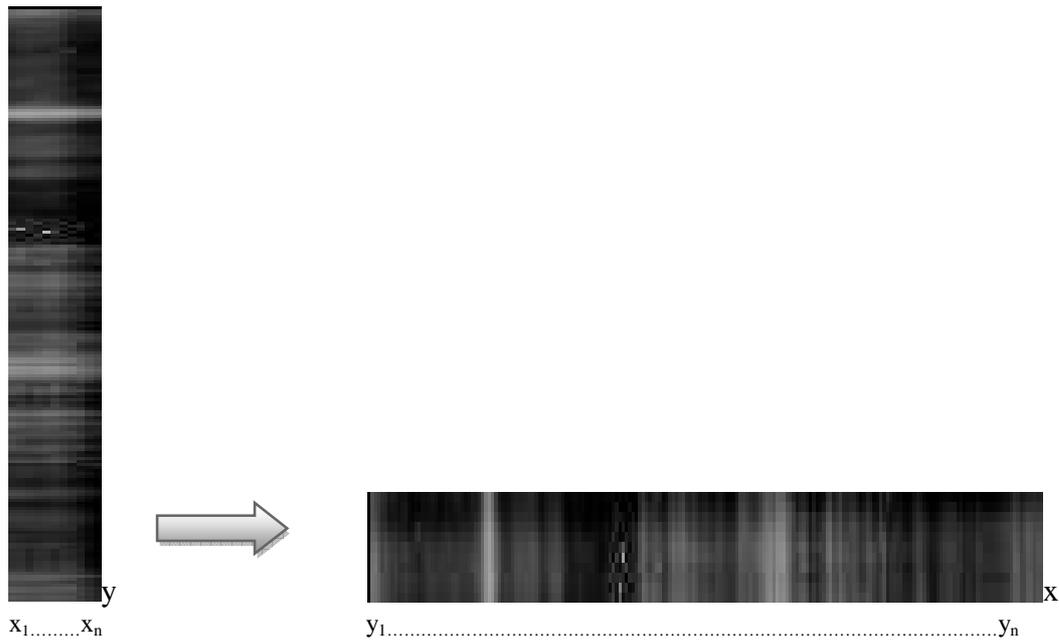


Figura 33: Carril Girado horizontalmente a partir de la imagen en A.

Una vez girado el carril con valores es posible aplicarle la función SUM, extraída desde las librerías de Matlab, la cual tiene como propósito principal el poder generar un histograma que contenga los valores de la suma de pixeles por cada punto en el eje x, con el fin de poder visualizar en qué lugares del carril se concentran los niveles de luminiscencia más alto entregado por los valores cercanos al blanco. En base al carril visualizado en la Figura 33, se puede obtener el siguiente histograma de la suma de pixeles, representado en la Figura 34.

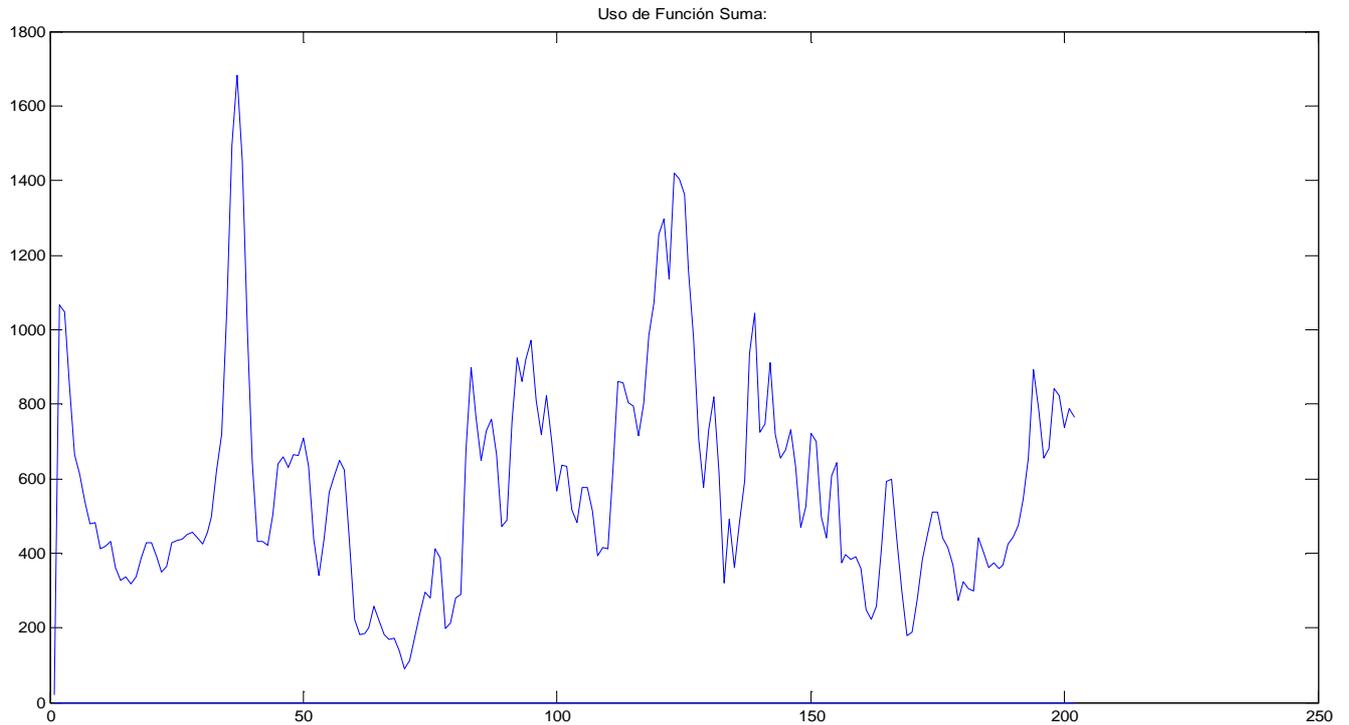


Figura 34: Histograma correspondiente a la suma de pixeles.

Desde la perspectiva de la programación la manera de llegar a éste histograma visto en la Figura 34 es a través de los siguientes códigos:

```

    gel = A
    gel = gel'
    [ny nx] = size(gel)
    gels = sum(gel)

```

En donde:

gel = A: *Gel* es la variable que almacenaba al carril *A*, obtenido desde los parámetros proporcionados por el programa principal.

gel = gel': Una vez que se haya almacenado el carril en la variable *gel*, este es girado horizontalmente en el momento de haber colocado en la parte superior del lado derecho una comilla simple (*gel'*).

[ny nx] = size(gel): Al momento de girar el carril (*gel*), es aplicado sobre él la función que almacena el tamaño correspondiente a la altura y anchura del carril.

$gelS = sum(gel)$: Se genera una nueva variable llamada $gelS$ la cual contiene los valores de las sumas de los píxeles de todas las columnas de la matriz $[ny \ nx]$, a través del uso de función SUM .

Después de aplicar la función SUM , se puede observar que en la mayoría de las cumbres del histograma de la Figura 34, existe mucha presencia de ruido, el cual es representado por las curvas con quiebres. Hecho que puede ser demostrado en la Figura 35, a través de los círculos sobrepuestos en cada una de ellas.

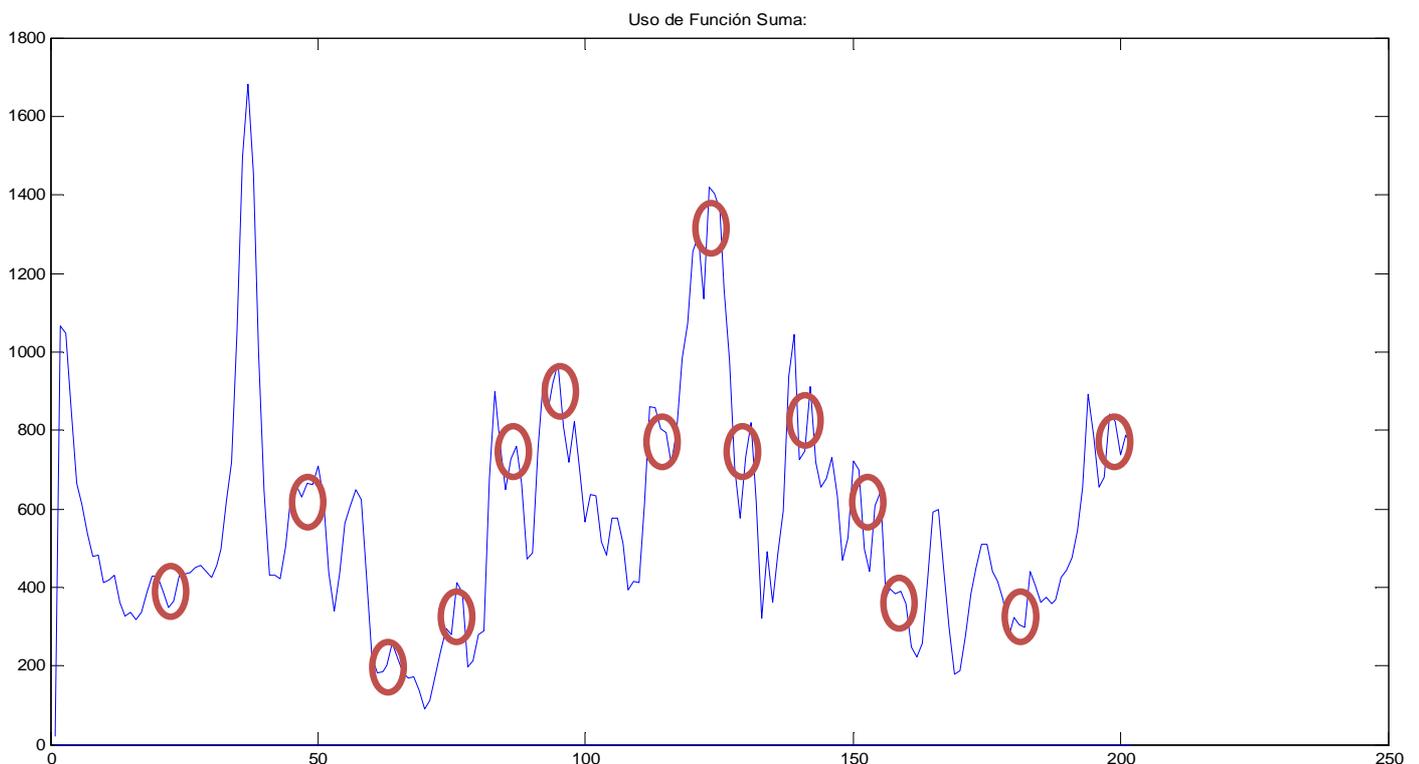


Figura 35: Demarcación de los principales ruidos presentados en el Histograma de suma de píxeles.

Para disminuir estas cumbres provistas de ruido, se debió crear un método de suavizamiento, el cual se encarga de eliminar los valores de intensidades que se encuentran por debajo del 25% de los cuartiles de intensidades. Para lo cual se ha tomado un valor constante igual a 200 en base a experimentos realizados por el Dr. Luis Ruedas¹⁴ en la

¹⁴ Ver Anexo A: Band Detection in Random Amplified Polymorphism DNA Images” del Dr. Luis Ruedas.

Universidad de Concepción, permitiendo con ello descartar aquellos valores que no son de importancia en la detección de las bandas. Permitiendo con ello la obtención del siguiente histograma de suavizamiento en la Figura 36.

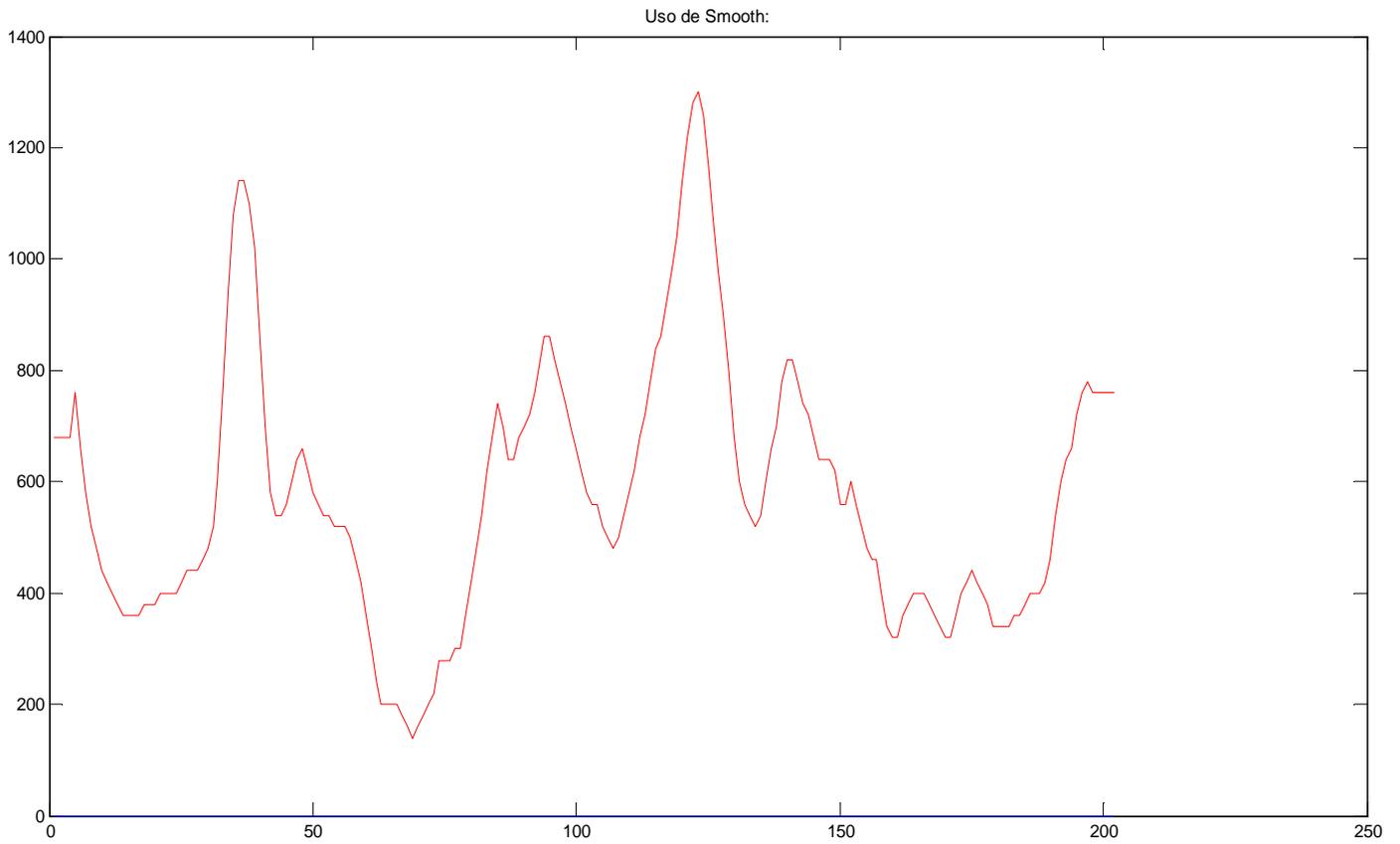


Figura 36: Histograma de la suma de pixeles suavizado.

Paso 5: Normalización del Histograma de Suavizamiento

Una vez obtenida la curva suavizada del histograma de suma de pixeles, se da paso a su Normalización, la cual funciona a través de derivadas parciales sobre los puntos mínimos de cada cumbre, llevándolos a una misma base o nivel cero, como se aprecia en la Figura 37, en donde se muestra la curva de color azul, la correspondiente a la de suavizamiento, y la negra la de normalización. Permitiendo con ello mejorar la detección de las bandas y por ende la eliminación del ruido.

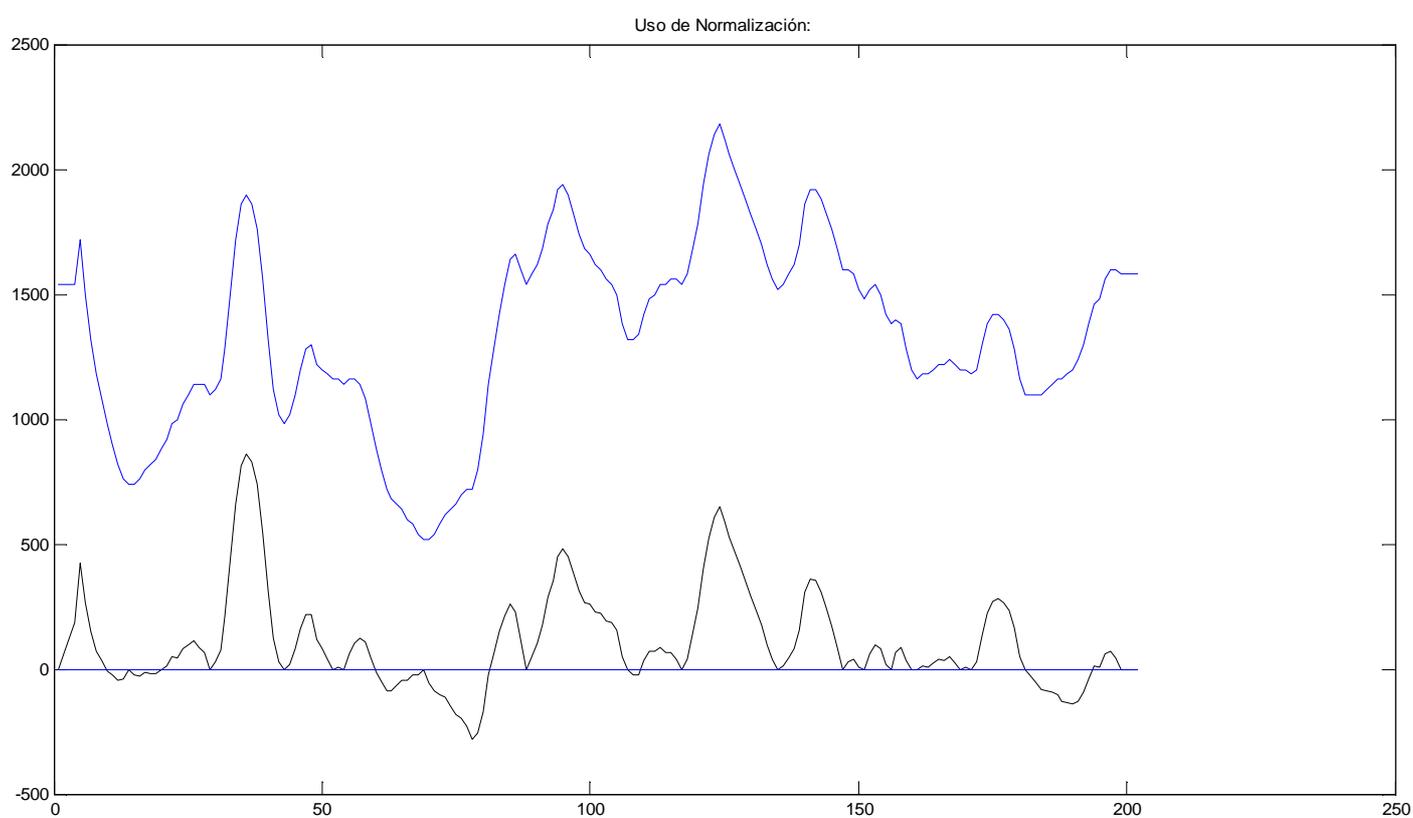


Figura 37: Histograma de Normalización

Paso 6: Búsqueda de las posiciones de cada Banda

Una vez que se haya efectuado completamente la normalización al carril, se da paso a la búsqueda de las posiciones de cada banda.

Proceso que comenzará con el recorrido de todo el largo del carril y en base al histograma de normalización, detectando en él las cumbres y los puntos más altos en ella, puntos que serán almacenados en un arreglo denominado `bands[]`, sin embargo antes de guardar dicha posición se verificará que el valor correspondiente a cada uno de sus píxeles será igual o cercano al 255, ya que éste representa al color blanco o al más provisto de luminiscencia. Una vez que se haya confirmado la banda, será marcada con una línea de color negro de la misma manera como se aprecia en la Figura 38.

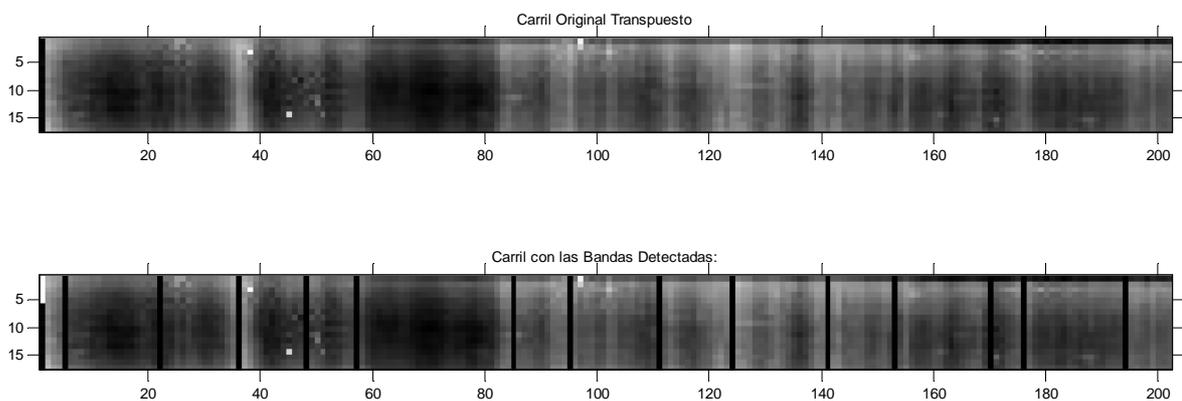


Figura 38: Bandas Identificadas y Marcadas

Originando con ello el siguiente vector denominado *Bands*, el cual contiene las posiciones de cada banda, posiciones que además pueden ser rectificadas por el lector con la Figura 40.

bands

[5 22 36 48 57 85 95 111 124 141 153 170 176 194]

Este mismo proceso se puede utilizar para la identificación y detección de los demás carriles, lo cual se puede representar en la Figura 39 y 40, en donde la primera muestra a la imagen original y la segunda el conjunto de carriles extraídos con sus bandas respectivamente marcadas.

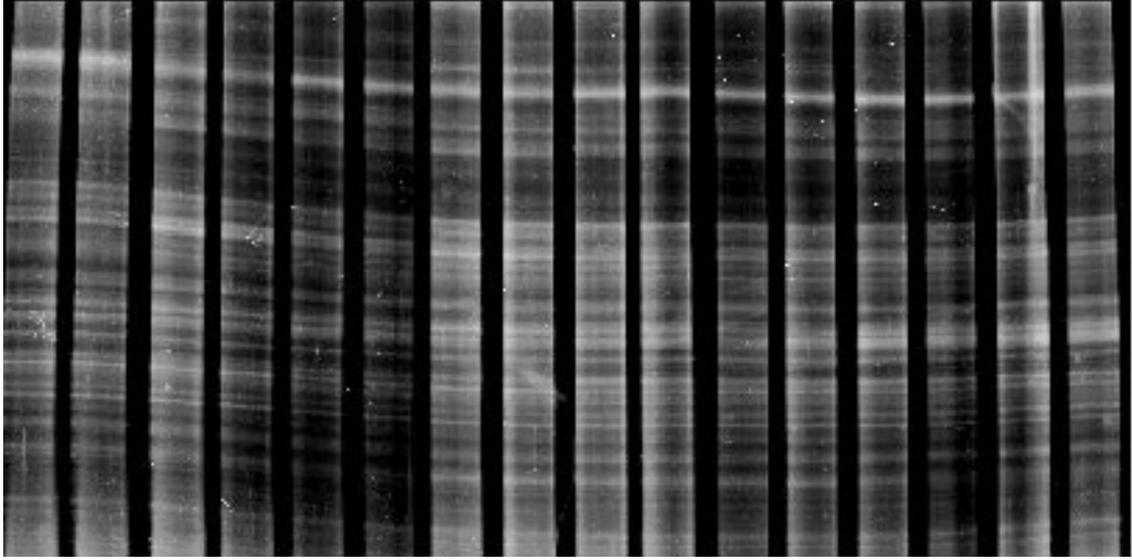
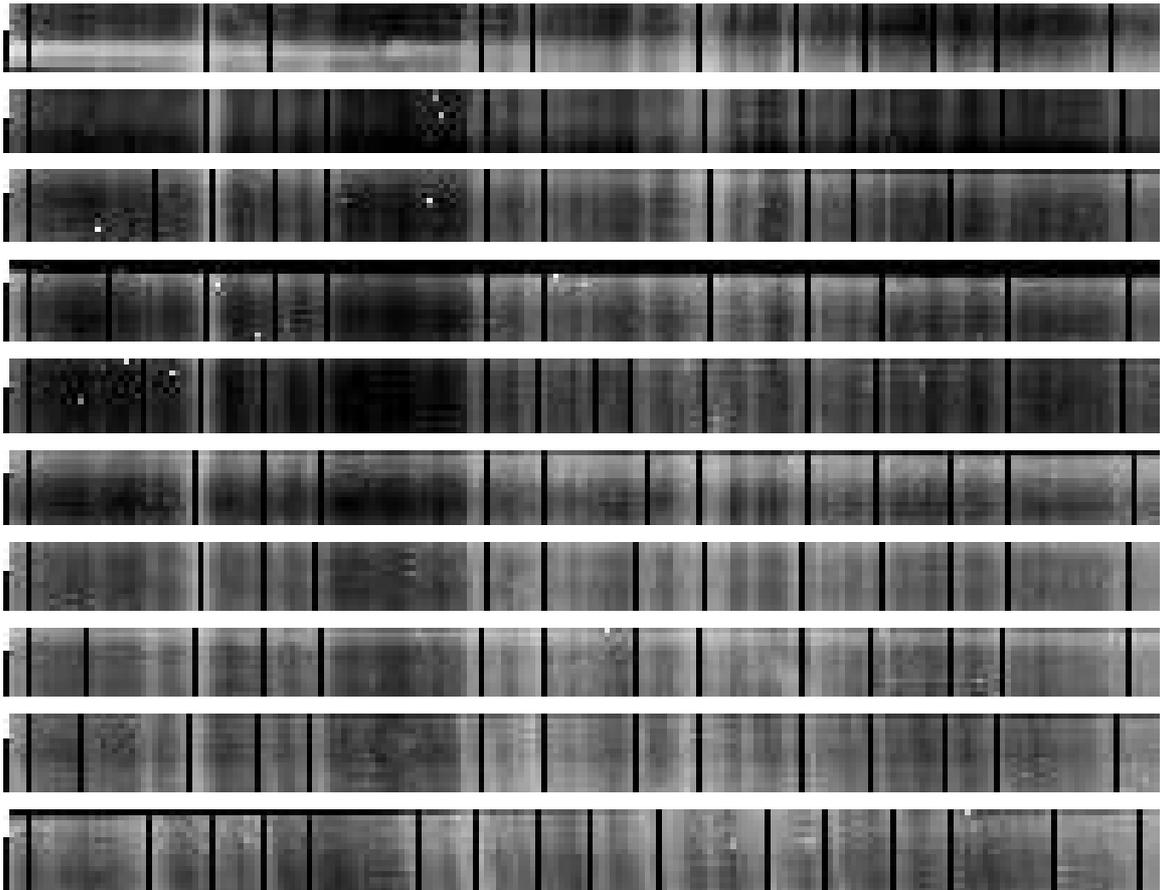


Figura 39: Imagen DGGE de prueba



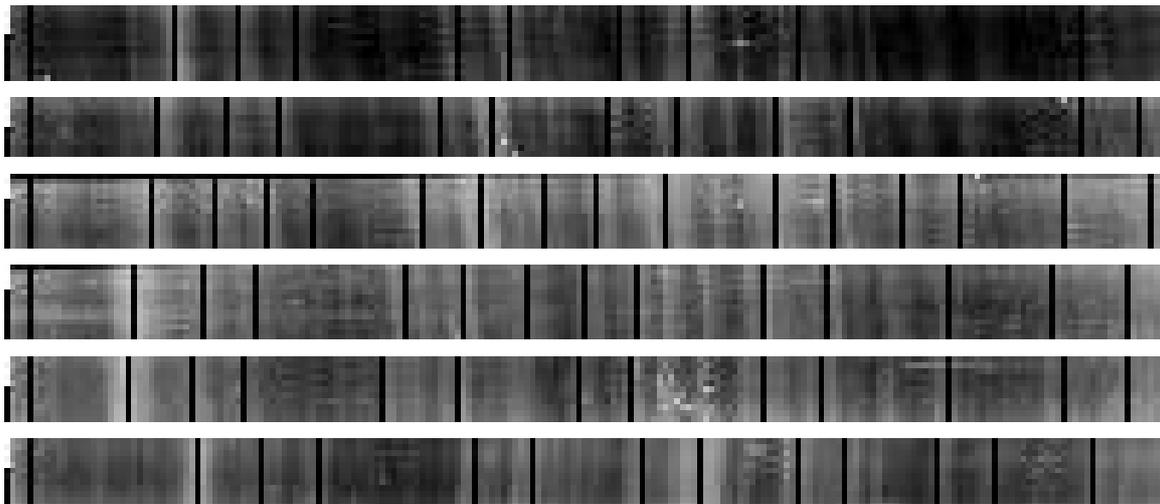


Figura 40: Carriles con las Bandas Identificadas.

Una vez que se haya finalizado la etapa concerniente a la Detección e Identificación de Bandas para cada carril de la Figura 40. Estos son unidos o concatenados, dando origen a una nueva imagen como se aprecia en la Figura 41, como lo ha sugerido el usuario principal, debido a que de esta manera se puede determinar la similitud, la riqueza o la diversidad de especies, pues al realizar esta concatenación se puede inferir sobre la familiaridad existente entre los carriles a partir de las Bandas marcadas en él.

Esta concatenación descrita anteriormente es realizada a partir de las siguientes líneas de código:

```
B1 = cat(2, C1',C2');
B2 = cat(2, C3',C4');
B3 = cat(2, C5',C6');
B4 = cat(2, C7', C8');
B5 = cat(2,C9',C10')

D1 = cat(2, B1, B2);
D2 = cat(2, B3, B4);
D3 = cat(2, D1, D2);
D4 = cat(2, D3, B5);

imwrite(D4,strcat(fileName12,'CARRILES_CONCATENADOS.jpg'),'JPEG');
```

En donde:

$B1 = cat(2, C1', C2')$: La variable $B1$, almacena la concatenación resultante entre los carriles $C1$ y $C2$ respectivamente, a través de la función cat . Sentencia que se repite hasta el último par de carriles, que en este ejemplo sería $C9$ y $C10$.

$D1 = cat(2, B1, B2)$: La variable $D1$, almacena la concatenación resultante entre los pares de carriles ya concatenados $B1$ y $B2$ respectivamente, a través de la función cat .

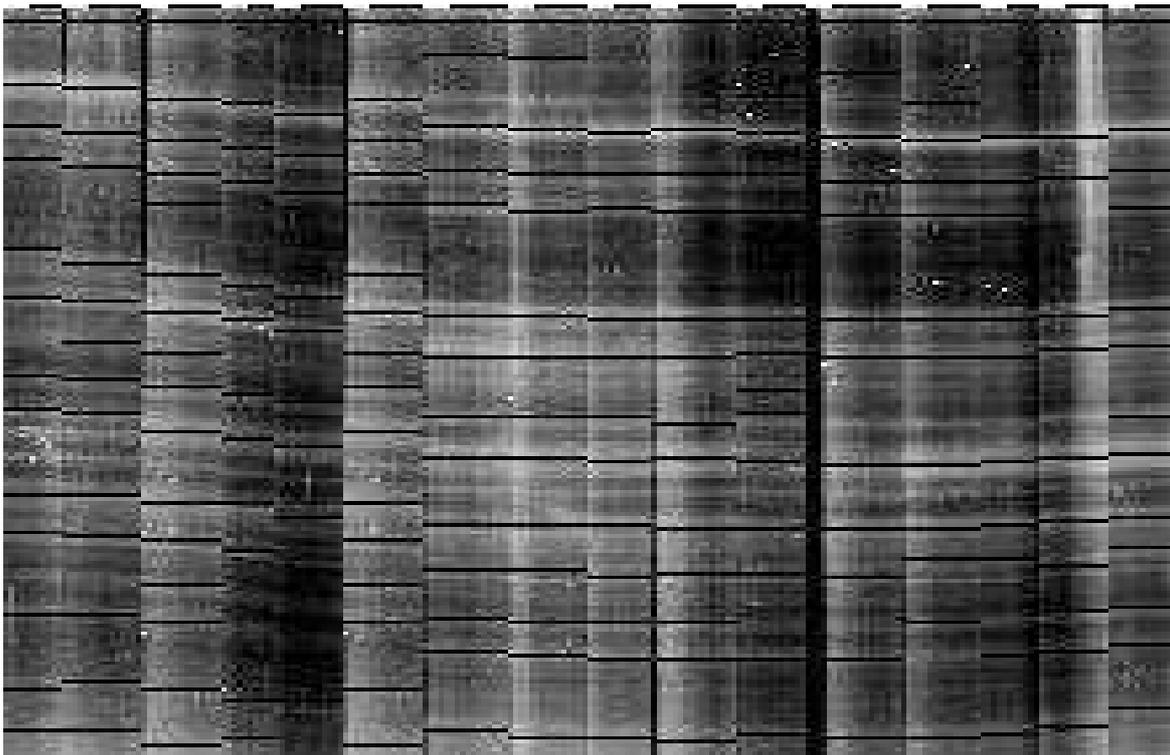
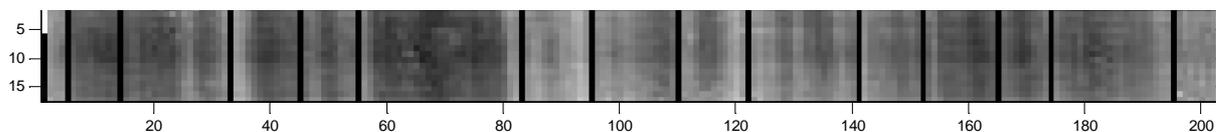


Figura 41: Concatenación de los Carriles Extraídos de la Figura 37.

Para lo cual se ha hecho uso nuevamente de los vectores *Bands*, ya que es la base para afirmar sobre el supuesto de familiaridad. Por ejemplo en los siguientes carriles las posiciones de las bandas parecieran estar en la misma posición, sin embargo ésta deducción “pareciera” es llevado a cabo de manera visual, porque si comprobamos los vectores *bands* correspondiente a cada carril, podríamos afirmar o rechazar esa hipótesis.

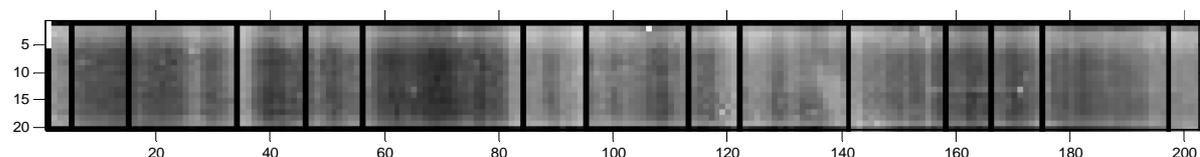
Carril 7 de la Figura 41



Bands

[5 14 33 45 55 83 95 110 122 141 152 165 174 195]

Carril 8 de la Figura 41



Bands

[5 15 34 46 56 84 95 113 122 141 158 166 175 197]

En conclusión se puede observar que las bandas número 1, 7, 9 y 10 en ambos carriles están situadas en las mismas posiciones, lo cual podría indicar la similitud o familiaridad que existe entre las especies encontradas en ellos, afirmando con ello que las muestras extraídas en el mar de la cuarta región actúan de la misma forma que la extraída de un mar de la octava región. Concluyendo además que este módulo no solo permite saber la riqueza de especies que se encuentran en cada carril, sino que además estima la similitud entre ellos.

Cabe señalar además, que los resultados obtenidos por este módulo serán detallados y comparados con los arrojados por el programa Quantity-One, en la sección 4.7.

4.4. Desarrollo del Sub-Módulo CDB

Una vez que se haya terminado de desarrollar el módulo DIDB, se puede dar paso a la cuantificación de las Bandas detectadas a través de los siguientes pasos, los cuales se resume en el siguiente diagrama de flujo representado en la Figura 42.

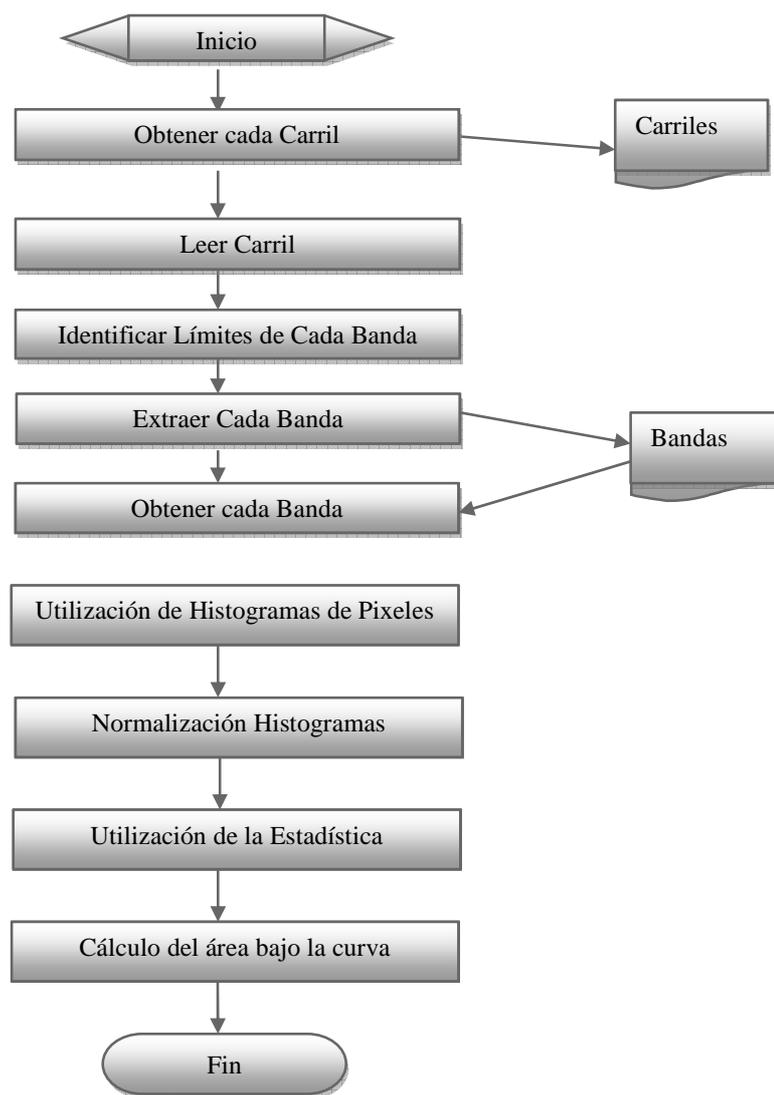


Figura 42: Diagrama de Flujo del Módulo CDB

Paso 1: Obtención y lectura de cada Carril

El primer paso antes de comenzar con el desarrollo de este módulo, corresponde a la obtención de la imagen, que en este caso sería la lectura de cada carril por separado. Paso muy similar al realizado en la captura de la imagen DGGE, ya que la única diferencia radica en la ruta del directorio, en donde se encuentran todos los carriles obtenidos de la imagen DGGE original.

```
fileName = 'IMAGENES_DGGE/DGGE1/CARRILES/1';
imageName = strcat(fileName, '.jpeg');
A = imread(imageName);
```

En donde:

fileName: Corresponde a la ruta del directorio en donde está almacenada el carril seleccionado.

imageName: Realiza la conversión con la ayuda de la función *strcat* (función Matlab) la concatenación de la ruta con el formato en el que ha sido guardado, el cual puede ser TIF o JPEG.

A: Es la variables tipo matriz que almacena al primer carril, a través del comando *imread*.

Paso 2: Identificación de los límites de cada Banda

El paso posterior a la obtención del carril, corresponde a la identificación de los límites superiores e inferiores de cada Banda, con el fin de poder extraer cada una de ellas y así facilitar la labor concerniente a su cuantificación. Esta labor se realiza mediante el uso de derivadas parciales para la detección de bordes, ya que permiten representar de mejor manera los cambios bruscos de tonalidades entre los pixeles con valores cercanos al negro y blanco (0 y 255), la cual opera sobre los valores proporcionados por la función SUM, la cual se encarga de leer pixeles por columnas. Para lo cual el carril capturado debe de estar en posición horizontal, como se muestra en la Figura 43.

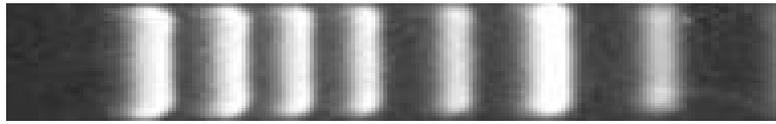


Figura 43: Carril obtenido para la detección de los límites de las bandas.

Para ellos ha sido necesario, crear una función que permita realizar la búsqueda de los límites superiores e inferiores para cada banda, a través de *función_cdb.m*, la cual recibe como parámetros, el carril en estudio y las posiciones de las bandas entregadas por la variable global *bands*, a partir del módulo DIDB.

Esta función además de necesitar de estos parámetros, hace uso de la función *normalize_hist_band.m*, ya que ésta sub-función nos permite obtener todas las posiciones en donde su valor es igual a cero, debido a que el 0 como se dijo anteriormente representa a la ausencia de color o de luminosidad, entregando con ello un vector que contendrá a aquellas posiciones denominado *bordes*, sin embargo no todos los valores contenidos en el vector corresponden a los límites reales de cada banda, para lo cual estos deben de ser validados y guardados finalmente en dos nuevos vectores denominados: *borde_izquierdo* y *borde_derecho*, ya que en base a estos valores se podrá dar paso a la extracción de las bandas.

Paso 3: Extracción de Cada Banda

Para realizar la extracción o segmentación de cada banda, se debió de crear una función llamada *función_extrae_bandas.m*. La cual permite como lo dice su nombre, extraer cada banda a partir del uso de *función_borde_bandas.m*, debido a que esta función le permitía obtener las posiciones de los límites, a través de los vectores *borde_izquierdo* y *borde_derecho*. Permitiendo con ello, facilitar la extracción y poder guardar todas las bandas de cada carril en un nuevo fichero o archivo, para poder ser procesada en un paso posterior a éste.

Desde la perspectiva del código, lo descrito anteriormente puede ser representado de la siguiente manera:

```
function [bandas] = funcion_extrae_bandas(A,borde_izquierdo,
borde_derecho)
global borde_izquierdo
global borde_derecho

%SE ENCARGA DE GUARDAR CADA BANDA SEGMENTADA

fileNameBandas11 = 'MAS_DGGE/DGGE5/CARRIL11/';

banda11_1 = A'(:,borde_izquierdo11(1,1):borde_derecho11(1,1));
imwrite(banda11_1, strcat(fileNameBandas11, 'banda1.jpg'), 'JPEG');
```

En donde:

function [bandas] = funcion_extrae_bandas(A,borde_izquierdo,borde_derecho): Declara a *función_extrae_bandas* como una función a través de *function*, la cual recibe como parámetros al carril *A*, y los vectores que almacenan todos los límites inferiores y superiores de todas las bandas de ese carril

fileNameBandas11 = 'MAS_DGGE/DGGE5/CARRIL11': Es la ruta del directorio en donde serán almacenadas las bandas que serán extraídas.

banda11_1 = A'(:,borde_izquierdo11(1,1):borde_derecho11(1,1)):: Variable que almacenará a la primera banda segmentada.

imwrite(banda11_1, strcat(fileNameBandas11, 'banda1.jpg'), 'JPEG'): Permite guardar la banda la primera banda segmentada con el nombre de *banda1.jpg*.

Un ejemplo de segmentación se puede apreciar en la Figura 44, la que corresponde a la primera banda delimitada en la Figura 43.

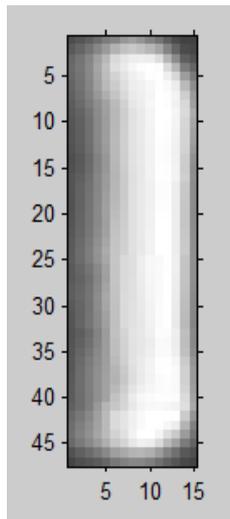


Figura 44: Banda 1 Segmentada

Paso 4: Utilización de Histograma de Suma de Píxeles

Cuando se haya terminado todo el proceso de segmentación de bandas y hayan sido guardadas en un archivo en común, el paso que continúa esta determinado por la lectura de cada una de ellas, con el fin de poder procesarlas a través del uso de Histogramas de Suma de Píxeles, como se observa en la Figura 45, el cual representa a las intensidades de luminiscencia de la primera banda.

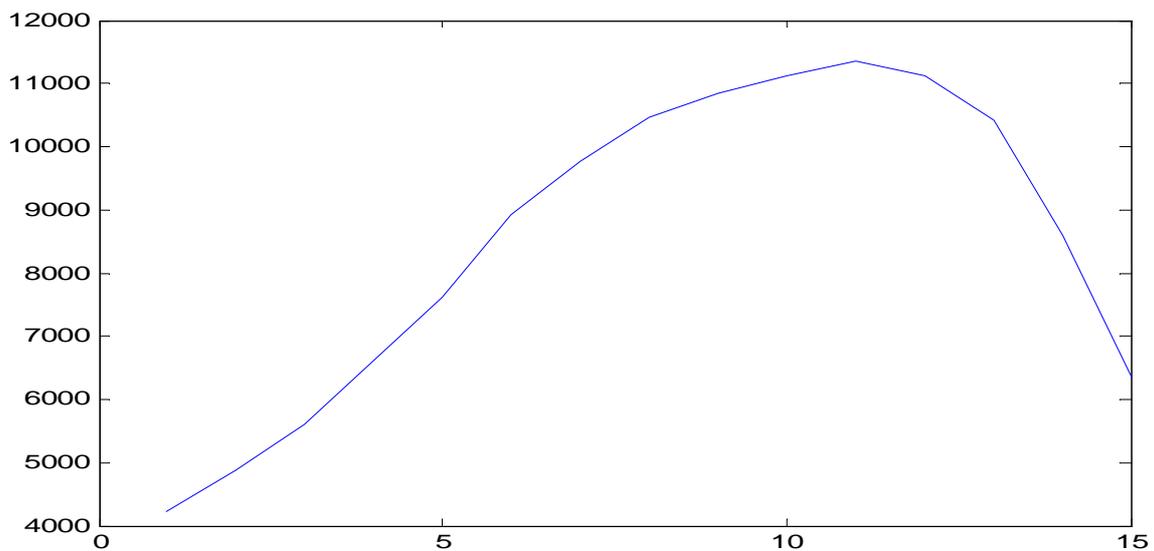


Figura 45: Histograma de la Suma de Píxeles de la Banda 1.

La importancia del uso de histogramas en esta sección, radica principalmente en las utilidades que brinda para el desarrollo de la alternativa de cuantificación que se está abordando, la cual tiene estrecha relación con las intensidades de luminiscencia, y los valores entregados por la función *SUM*, la cual permite realizar la sumatoria de los pixeles por cada columna. Conllevando con ello a la realización del cálculo del área que se encuentra bajo la curva *SUM*. Curva que necesita aún del manejo de filtros que permitan suavizarla, eliminando con ello todo ruido o información irrelevante, de la misma manera cómo se utilizó en el módulo DIDB. Obteniendo entonces la siguiente curva de suavizamiento demostrada en la Figura 46, la que toma como base el histograma de suma de pixeles visto anteriormente.

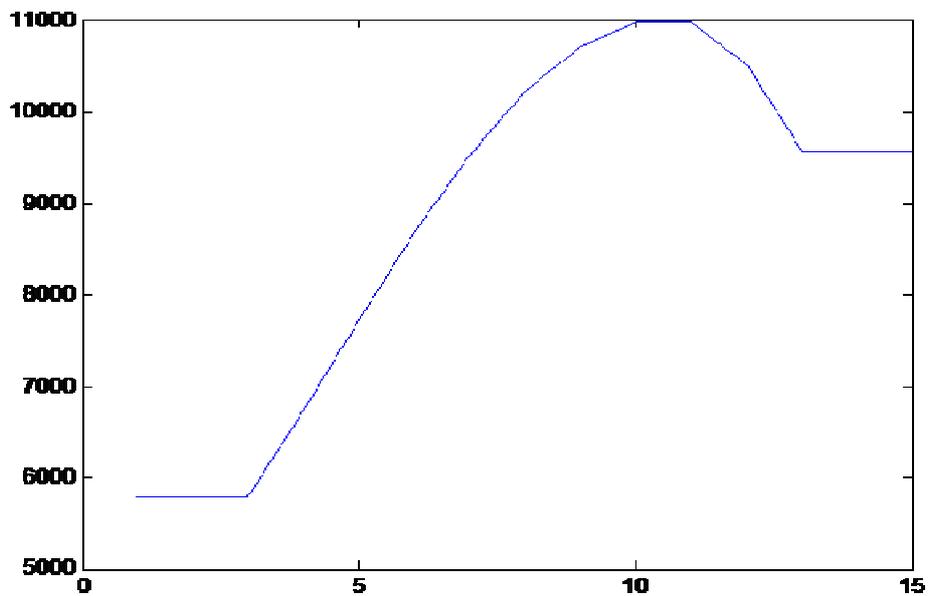


Figura 46: Curva de Suavizamiento del Histograma de Suma de Pixeles

Paso 5: Normalización de Histogramas

Como se pudo apreciar en el histograma de suavizamiento, los puntos de los valles están en diferentes posiciones con respecto al eje y, lo que dificulta la tarea del cálculo del área bajo la curva para encontrar los puntos bases de cada límite, vale decir que estos puntos deben de ser $(x_1, 0)$, $(x_2, 0)$, para lo cual es necesario aplicar la funcionalidad que permita bajar la curva hacia un mismo nivel, para lo cual los puntos mínimo son llevados a un mismo nivel gracias a la función *normalizeHist_Band.m*, la cual opera de la misma manera que la normalización de los carriles. Arrojando con ello el siguiente histograma de normalización visualizado gráficamente en la Figura 47.

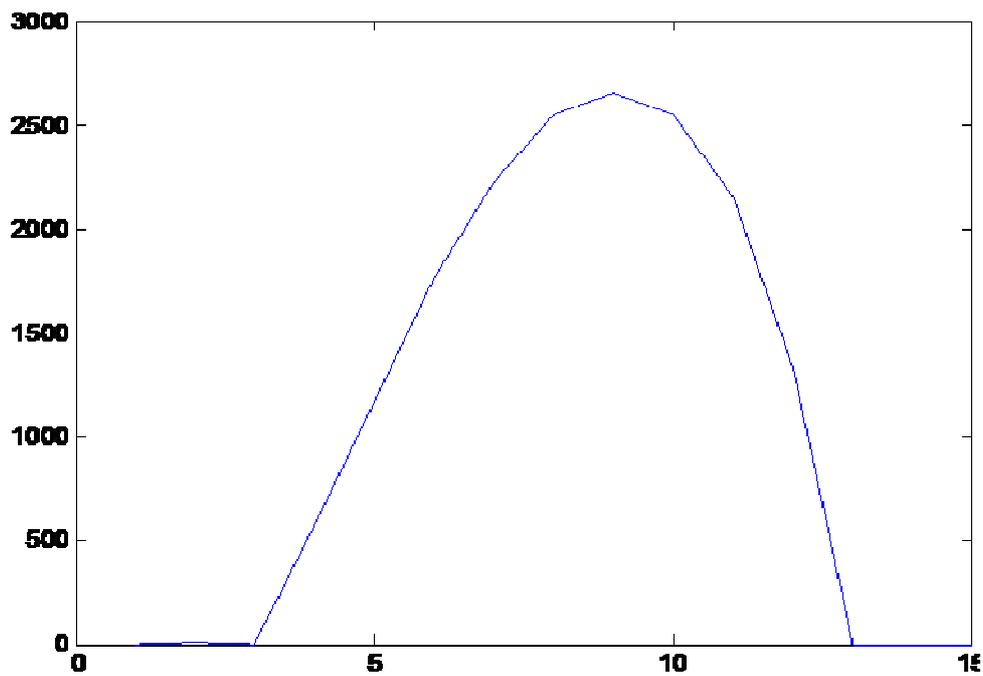


Figura 47: Histograma de Normalización

Paso 6: Utilización de la Estadística

Como se ha señalado en capítulos anteriores, la solución propuesta en el desarrollo del módulo CDB, consiste en utilizar algunos conceptos de la estadística, principalmente en el uso de la Distribución de la Normal, con el fin de poder entregar un valor numérico para la estimación de la densidad de las Secuencias de ADN provistas en cada banda.

En base al histograma de normalización se pudo obtener el cálculo de los siguientes estimadores estadísticos señalados en la Tabla 2.

Estimador	Valor
Media	1043
Varianza	1.2149e+006
Desviación	1.1022e+003

Tabla 2: Cálculo de los Estimadores Estadísticos.

Las fórmulas que se utilizaron para su cálculo se señalan a continuación:

Media

$$\mu = \frac{\sum_{i=1}^{ny} x_i}{n}$$

Donde:

- $\sum_{i=1}^{ny} x_i$

Corresponde a la sumatoria de todas las columnas de pixeles dada por la función SUM.

- ny

Corresponde al número total de las columnas de la banda (15).

Varianza

$$\sigma^2 = \frac{\sum_{i=1}^{ny} (x_i - \bar{x})^2}{ny}$$

Desviación Estándar

$$\sigma = \sqrt{\sigma^2}$$

Desde la perspectiva del código la forma en cómo se debería de calcular estos estimadores es la siguiente:

```
suma = sum(sum(C1));
media = suma/ny;
varianza = var(sum(C1));
desv_standar = sqrt(varianza);
```

Donde:

Suma: Realiza la labor de sumatoria, a través del uso de *sum* que permite sumar todo los pixeles por cada columna.

Var: Es una función de Matlab que permite obtener la varianza.

Sqrt: Permite obtener la raíz cuadrada de la varianza.

Paso 7: Cálculo del área bajo la curva

Después de haber calculado los estimadores estadísticos necesarios para el procesamiento de la Distribución Normal, se realiza la ejecución en base a la función NORMCDF del Toollbox de Estadísticas de MATLAB. Función por la cual permitirá obtener el siguiente valor, el cual correspondería en este caso a la Banda número uno del carril uno, en base a la sumatoria de los valores de la Tabla 3, lo cual cada X_i corresponde al valor de un punto (x,y) de la Banda. Por lo cual al realizar la sumatoria de todos los X_i , nos daría como resultado el siguiente valor numérico: **8.4771e+003**, valor que corresponde a la cuantificación de esa Banda.

X_i	Valor	X_i	Valor	X_i	Valor
1	0	6	0.8797	11	1.0785
2	0	7	1.1132	12	0.6650
3	0	8	1.2747	13	0
4	0.2810	9	1.3289	14	0
5	0.5822	10	1.2739	15	0

Tabla 3: Valores de la sumatoria de cada columna del eje X de la Banda 1

En la Tabla 4 se puede observar todos los valores asociados al carril en estudio, sobre las siete bandas detectadas en él. Valor que corresponde al área bajo la curva, proporcionado por la ayuda de la distribución normal, resultados además que están relacionados con la intensidad de pixeles.

Banda N°	Valor
1	8.4771e+003
2	4.6274e+003
3	3.4850e+003
4	1.1298e+004
5	1.0014e+004
6	1.4285e+004
7	1.1245e+004

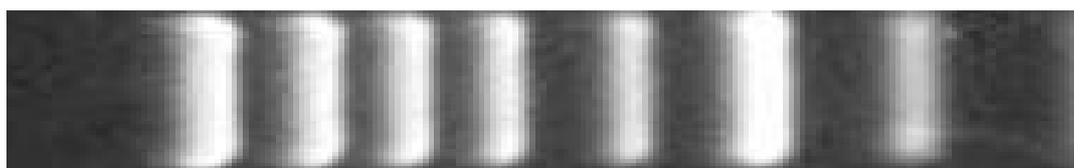
Tabla 4: Resumen de los Valores de Cuantificación

Una vez terminado el proceso de cuantificación por cada Banda, se realiza la sumatoria de todas las áreas correspondiente a cada una de ellas, con el sentido de ver la participación de cada Secuencia de ADN o Banda en el carril que se ha señalado como ejemplo en este módulo. Resultados que se resumen en la Tabla 5.

Banda N°	Valor	Participación	Porcentaje
1	8.4771e+003	0.1336	13,36 %
2	4.6274e+003	0.0730	7.3 %
3	3.4850e+003	0.0549	5,49 %
4	1.1298e+004	0.1781	17,81 %
5	1.0014e+004	0.1579	15,79 %
6	1.4285e+004	0.2252	22,52 %
7	1.1245e+004	0.1773	17,73 %
Total = 8	6.3432e+004	1	100%

Tabla 5: Resumen de la Participación de Cada Banda en el Carril.

Como se pudo apreciar en la Tabla 5, la mayor participación la tendría la Banda número 6, la cual además tiene una mayor concentración en sentido de intensidades en pixeles. Permitiendo con ello concluir en este ejemplo que su densidad molecular es mayor que las demás bandas, como se puede observar de manera gráfica en el siguiente carril, que sirvió de ejemplo para la demostración de cómo opera el módulo CDB.



Sin embargo, aún queda mucho camino para llegar a una alternativa perfecta con respecto a la cuantificación, debido a que han habido varios conflictos con respecto a la detección de bordes, lo cual dificulta la extracción de las bandas y por ende los resultados su cuantificación. Debido a algún problema en el algoritmo que realiza la búsqueda de las posiciones de cada límite.

4.5. Tecnologías Utilizadas

La tecnología utilizada en el desarrollo del proyecto es el entorno Matlab®. Este entorno de trabajo, ofrece una gama de beneficios en el desarrollo de aplicaciones computacionales, que permiten el desarrollo de proyectos como éste. MATLAB [14] integra análisis numérico, cálculo matricial, procesamiento de señales y visualización gráfica en un entorno completo donde los problemas.

MATLAB dispone en la actualidad de una gama de programas de apoyo especializados, denominados Toolboxes, que extienden significativamente el número de funciones incorporadas en el programa principal. Estos Toolboxes cubren en la actualidad prácticamente casi todas las áreas principales en el mundo de la ingeniería y la simulación, destacando entre ellos el 'toolbox' de proceso de imágenes, señal, control robusto, estadística, análisis financiero, matemáticas simbólicas, redes neurales, lógica difusa, identificación de sistemas, simulación de sistemas dinámicos, etc. es un entorno de cálculo técnico, que se ha convertido en estándar de la industria, con capacidades no superadas en computación y visualización numérica.

Características:

- Alto rendimiento en cálculos numéricos.
- Posee un nivel avanzado en visualizaciones gráficas.
- El lenguaje utilizado es de alto nivel, el cual se fundamenta en la utilización de vectores, matrices y arreglos.
- Posee un conjunto de funciones de aplicaciones de alto nivel.

Las herramientas o librerías de implementación, vienen incorporadas dentro de la plataforma MATLAB, siendo las más usadas en el proyecto:

- Image Processing Toolbox.
- Statics Processing Toolbox.

Image Processing Toolbox.

Este Toolbox proporciona a MATLAB de un conjunto de funciones que amplía las capacidades del producto para realizar desarrollo de aplicaciones y de nuevos algoritmos en el campo del proceso y análisis de imágenes. El entorno matemático y de creación de MATLAB es ideal para el procesamiento de imágenes, ya que estas imágenes son, al fin y al cabo, matrices. Este toolbox incorpora funciones para:

- Diseño de filtros.
- Mejora y retocado de imágenes.
- Análisis y estadística de imágenes.
- Operaciones morfológicas, geométricas y de color.
- Transformaciones 2D.

El proceso de imágenes es un campo de trabajo absolutamente crucial para aquellos colectivos e industrias que estén trabajando en áreas como diagnóstico médico, astronomía, geofísica, ciencias medioambientales, análisis de datos en laboratorios, inspección industrial, etc. Los programas actuales de procesamiento y análisis de imágenes se clasifican actualmente en dos categorías: librerías de bajo nivel para programadores profesionales y paquetes de aplicación con capacidades limitadas de personalización. Ambos tipos de aplicaciones están, generalmente, pensados para tareas básicas de visualización de datos 'dering'. Sin embargo, muchos de ellos adolecen de la posibilidad de efectuar análisis numéricos de los mismos. El Image Processing Toolbox entra dentro de la categoría de familias de funciones que, desde el entorno de trabajo de MATLAB , permitirá al profesional efectuar una exploración exhaustiva y desde un punto de vista matemático de las imágenes y gráficos que se deseen tratar o analizar.

Algunas de las funciones más importantes incluidas dentro de este toolbox son las siguientes:

1. Análisis de imágenes y estadística.
2. Diseño de filtros y recuperación de imágenes.
3. Mejora de imágenes.

4. Operaciones morfológicas.
5. Definición de mapas de colores y modificación gráfica.
6. Operaciones geométricas.
7. Transformación de imágenes.
8. Proceso de bloques

Listado de las Funciones utilizadas del Image Processing Toolbox

1. **Función IMREAD:** Esta función permite leer desde archivo una imagen.
2. **Función RGB2GRAY:** Crea una imagen de intensidad en escala de gris a partir de una imagen RGB.
3. **Función IMWRITE:** Permite guardar una imagen en un archivo.
4. **Función IMSHOW:** Permite mostrar por pantalla una imagen.
5. **Función SIZE:** Permite conocer el tamaño de una imagen de orden $N \times M$.
6. **Función IMROTATE:** Permite girar una imagen en base a un ángulo determinado.

Cabe destacar que todas las funciones vistas en esta sección se detallaran en el Anexo D de este documento.

Statics Processing Toolbox.

La utilización de este paquete de herramientas de MATLAB, facilita la tarea de poder cuantificar cada Banda previamente identificada y segmentada. Todo ello en base a la utilización de los Histogramas de Normalización. Para lo cual se tiene que calcular cada estimador estadístico, con el fin de obtener la Distribución Normal, debido a que es muy útil para poder calcular el área bajo la curva. Y con eso poder entregar un valor cuantitativo al usuario, expresado en intensidad de pixeles.

Listado de las Funciones utilizadas del Statics Processing Toolbox

1. **Función PDF:** Función de densidad de probabilidad (pdf) para una determinada distribución
2. **Función NORMCDF:** Función de la Distribución Normal

3. **Función VAR:** Permite calcular la varianza de una Banda previamente segmentada.
4. **Función MEAN:** La función Mean, permite calcular la media aritmética de un conjunto finito de valores.

Cabe destacar que todas las funciones vistas en esta sección se detallaran en el Anexo B de este documento.

4.6. Implementación del Sistema CYDB

Una vez que se hayan desarrollado ambos módulos, se puede dar paso a su implementación, la cual consiste en utilizar un conjunto de imágenes DGGE desde cualquier fuente, entiéndase por cualquier fuente, a las proporcionadas por el laboratorio o las encontradas en la web. Permitiendo con ello la generación de reportes y las comparaciones con el programa Quantity-One, en donde se profundizará además en el margen de error de ambos sistemas (CYDB y Q1), todo ello en base a mediciones estadísticas.

4.6.1. Pruebas en diferentes tipos de Imágenes DGGE en el módulo DIBD

El primer paso para su implementación del sistema CYDB, comienza con un conjunto de pruebas, en donde los resultados obtenidos de manera automática en ambos programas serán medidos mediante el uso de la estadística e histogramas.

Para llevar a cabo estas pruebas se hizo uso de imágenes que se encontraban en el laboratorio del Ceb-UDEC y algunas que fueron descargadas desde la web.

Prueba 1:

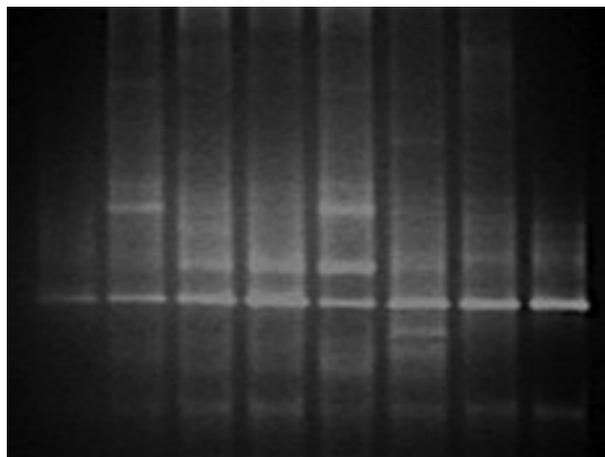


Figura 48: Imagen DGGE [5] para la prueba 1.

Carriles extraídos 8

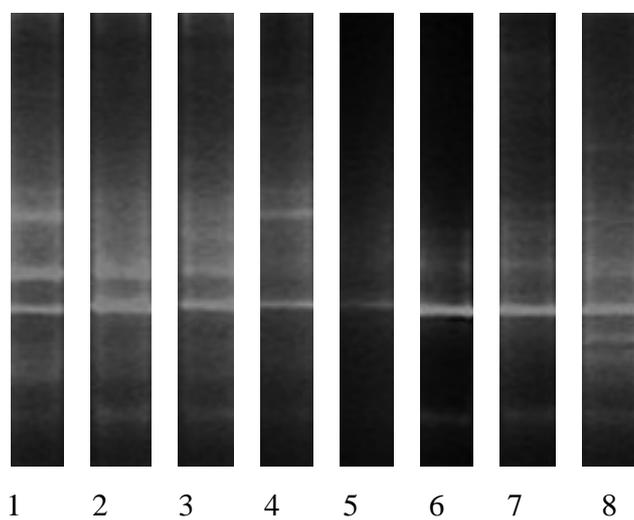


Figura 49: Carriles extraídos de la imagen DGGE de la Prueba 1.

Carril 1



- Uso del Módulo DIDB



- Uso del Quantity-One



Figura 50: Comparación de la detección de Bandas a través del módulo DIDB y Q1

Cabe señalar que los demás carriles se encuentran en el Anexo F.

En los siguientes recuadros señalados en las Tablas 6 y 7, se muestran las comparaciones resultantes sobre la detección automática de bandas en ambos programas. Además, en la Tabla 6 se agrega una columna, la cual representa las bandas reales, las que fueron ratificadas por un biólogo.

Carril	Bandas Detectadas con DIDB		Bandas Detectadas con Q1		Bandas Reales
	Correctas	Erróneas	Correctas	Erróneas	
1	2	0	1	7	2
2	3	0	2	6	3
3	4	0	4	4	4
4	3	0	3	7	3
5	5	1	4	5	6
6	6	0	6	5	6
7	4	1	4	4	5
8	3	1	3	7	3

Tabla 6: Cuadro comparativo de las Bandas encontradas en la Prueba 1.

Carril	Probabilidad DIDB		Probabilidad Q1	
	Éxito	Error	Éxito	Error
1	1	0	0,125	0,875
2	1	0	0,25	0,75
3	1	0	0,5	0,5
4	1	0	0,3	0,7
5	0,8	0,2	0,45	0,55
6	1	0	0,54	0,46
7	0,8	0,2	0,5	0,5
8	0,75	0,25	0,3	0,7

Tabla 7: Cuadro comparativo con las Probabilidades de Error y Éxito en cada programa.

La obtención de las probabilidad de éxito y error para cada programa, se realizó en base a los aciertos entre las bandas detectadas y las reales, las cuales como se menciono anteriormente fueron verificados por un biólogo del Ceb-UDEC. Lo que permitió generar los siguientes histogramas, los cuales reflejan gráficamente las diferencias notorias entre estos programas, dejando de manifiesto un margen de diferencia muy amplio.

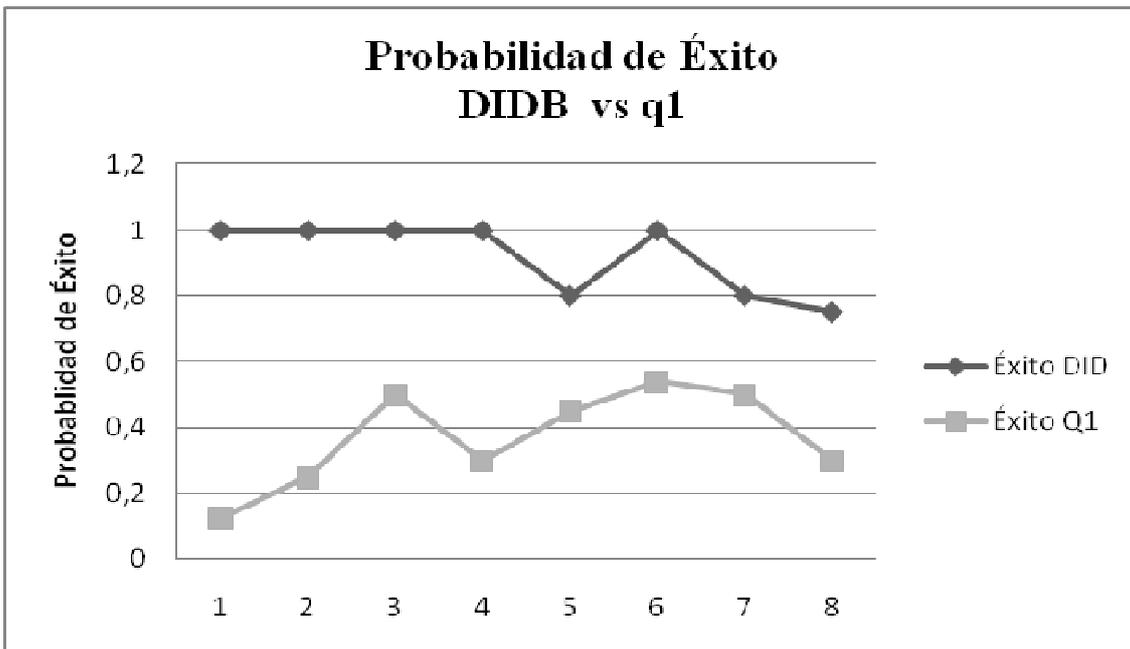


Figura 51: Gráfico comparativo entre el éxito en la detección en ambos programas.

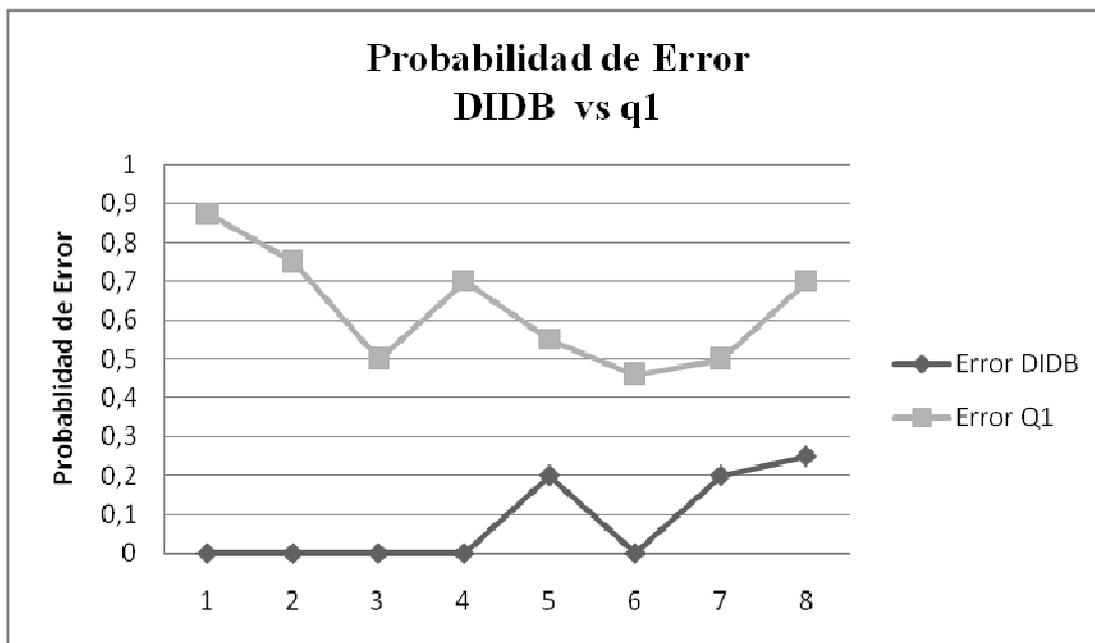


Figura 52: Gráfico comparativo entre el error en la detección en ambos programas.

Por lo tanto:

La Probabilidad de Éxito DIDB $0,91875 \approx 91,88\%$

La Probabilidad de Error DIDB $0,08125 \approx 8,125\%$

La Probabilidad de Éxito Q1 $0,303125 \approx 30,31\%$

La Probabilidad de Error Q1 $0,6294 \approx 62,94\%$

Lo cual deja de manifiesto la gran diferencia de ambas aplicaciones, en donde la ventaja competitiva la estaría llevando el módulo DIDB.

Prueba 2:

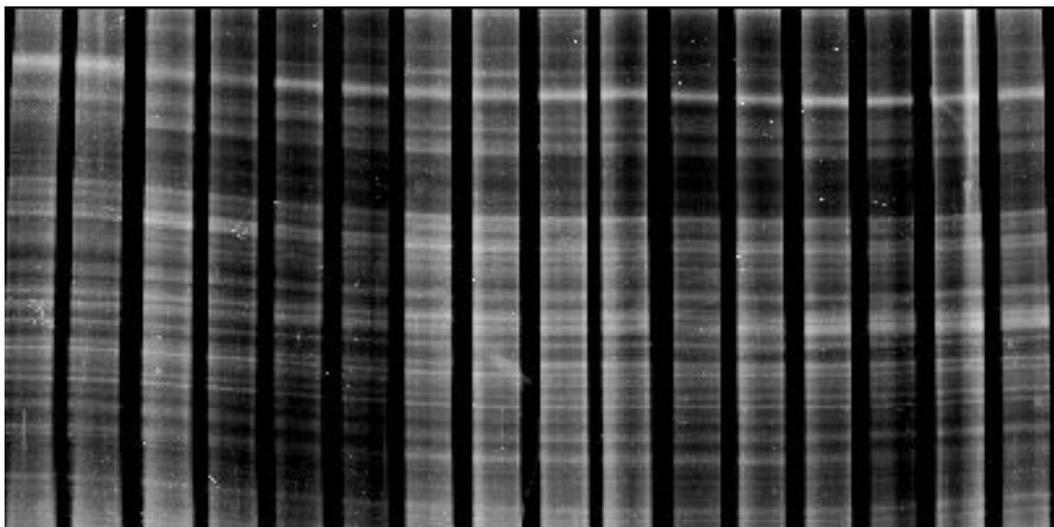


Figura 53: Imagen DGGE [6] para la prueba 2.

Carriles extraídos: 16

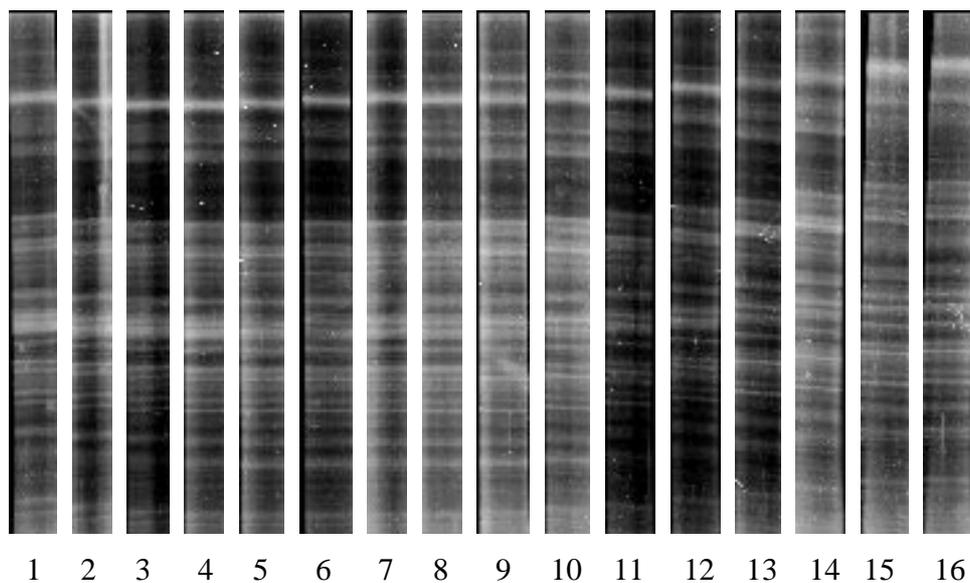


Figura 54: Carriles extraídos de la imagen DGGE de la Prueba 2

Al igual que en la prueba 1 el detalle de esta implementación se encuentra en el Anexo F.

En los siguientes recuadros señalados en la Tabla 8 y 9, se muestran las comparaciones resultantes sobre las detecciones automáticas de bandas por ambos programas. En la Tabla 8 se agrega una columna, la cual representa a las bandas reales ratificadas previamente por un biólogo.

Carril	Bandas Detectadas con DIDB	Bandas Detectadas con Q1	Bandas Reales
1	16	16	16
2	16	17	16
3	14	17	14
4	13	17	13
5	15	16	15
6	14	17	14
7	14	14	14
8	14	17	14
9	12	17	14
10	13	18	14
11	13	17	14
12	14	17	14
13	14	15	14
14	14	15	14
15	12	16	14
16	16	16	16

Tabla 8: Cuadro comparativo de las Bandas encontradas en la Prueba 2, a través de las dos aplicaciones

Cabe señalar que a pesar de que el número de bandas detectadas de manera automática por el Q1, este muy cercano al valor real varias de las bandas identificadas no corresponden por lo que el margen de error aumentaría considerablemente, como se muestra en la Tabla 9, la cual nos permite identificar la probabilidad de éxito, dada por las coincidencias entre las bandas reales con las detectadas por ambas aplicaciones y la probabilidad de error, dado por la falta de detección de una banda en particular, o por fallos en la detección, es decir, por detectar una banda que no existe. Esta última situación se puede reflejar en la columna *Erróneas y S/Detectarse*, de la sección del Q1, en donde 3+3 determina las bandas sin detectarse y las falsas.

Carril	Bandas Detectadas con DIDB		Bandas Detectadas con Q1		Bandas Reales
	Correctas	Erróneas y S/Detectarse	Correctas	Erróneas y S/Detectarse	
1	16	0	13	3 + 3	16
2	16	0	13	3 + 3	16
3	14	0	14	3	14
4	13	0	13	4	13
5	15	0	13	3+2	15
6	14	0	13	4+1	14
7	14	0	11	3+3	14
8	14	0	14	3	14
9	12	2	14	3	14
10	13	1	14	4	14
11	13	1	14	3	14
12	14	0	14	3	14
13	14	0	12	3+2	14
14	14	0	11	3+3	14
15	12	2	13	3+1	14
16	16	0	13	3+3	16

Tabla 9: Cuadro comparativo con las Bandas Detectadas de manera exitosa y con error.

Carril	Probabilidad DIDB		Probabilidad Q1	
	Éxito	Error	Éxito	Error
1	1	0	0,625	0,375
2	1	0	0,625	0,375
3	1	0	0,7857	0,214
4	1	0	0,693	0,307
5	1	0	0,67	0,33
6	1	0	0,643	0,357
7	1	0	0,571	0,429
8	1	0	0,7857	0,2143
9	0,857	0,143	0,7857	0,2143
10	0,929	0,071	0,714	0,286
11	0,929	0,071	0,7857	0,2143
12	1	0	0,7857	0,2143
13	1	0	0,6429	0,3571
14	1	0	0,571	0,429
15	0,857	0,143	0,714	0,286
16	1	0	0,625	0,375

Tabla 10: Cuadro comparativo con las Probabilidades de Error y Éxito en cada programa, para la prueba 2.

La obtención de las probabilidad de éxito y error para cada programa, se realizó en base a los aciertos entre las bandas detectadas y las reales, las cuales como se menciona anteriormente fueron verificados por un biólogo del Ceb-UDEC. Lo que permitió generar los siguientes histogramas, los cuales reflejan gráficamente las diferencias notorias entre estas aplicaciones, dejando de manifiesto un margen de diferencia muy amplio.

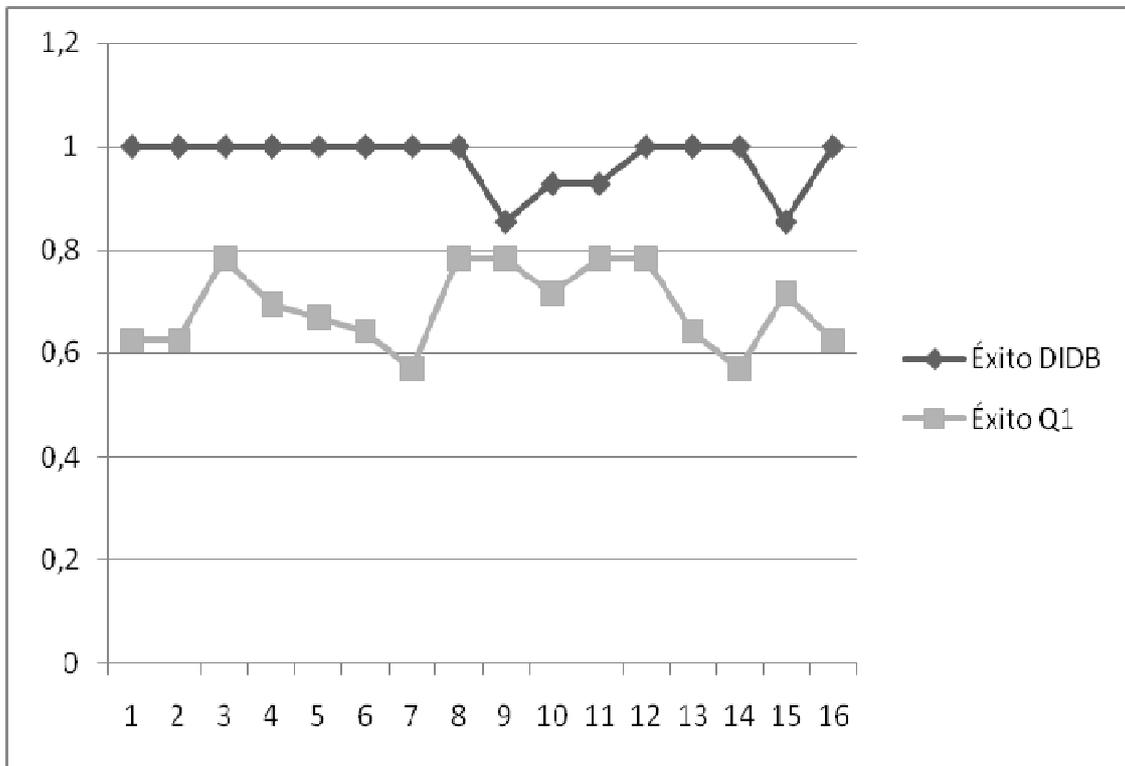


Figura 55: Gráfico comparativo entre el éxito en la detección en ambos programas.

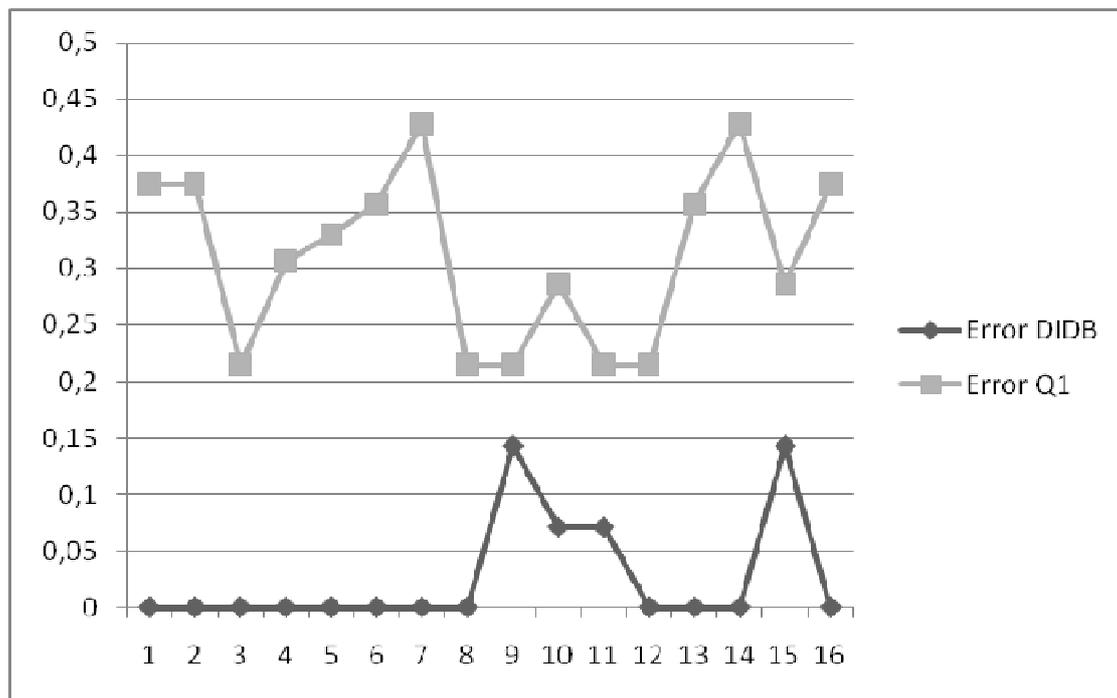


Figura 56: Gráfico comparativo entre el error en la detección en ambos programas.

Probabilidad de Éxito DIBB $0,97235 \approx 97,325\%$

Probabilidad de Error DIBB $0,02675 \approx 2,675\%$

Probabilidad de Éxito Q1 $0,6889 \approx 68,89\%$

Probabilidad de Error Q1 $0,311081 \approx 31,1081\%$

Prueba 3:

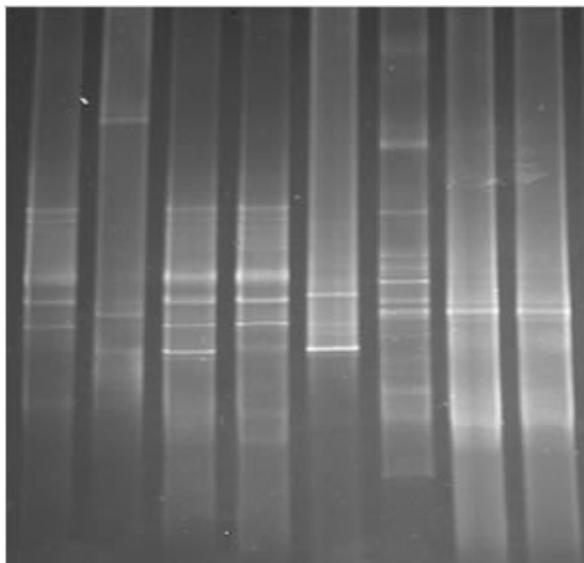


Figura 57: Imagen DGGE, proporcionada por el laboratorio del Ceb-UDEC, para la prueba 3.

Carriles extraídos: 8

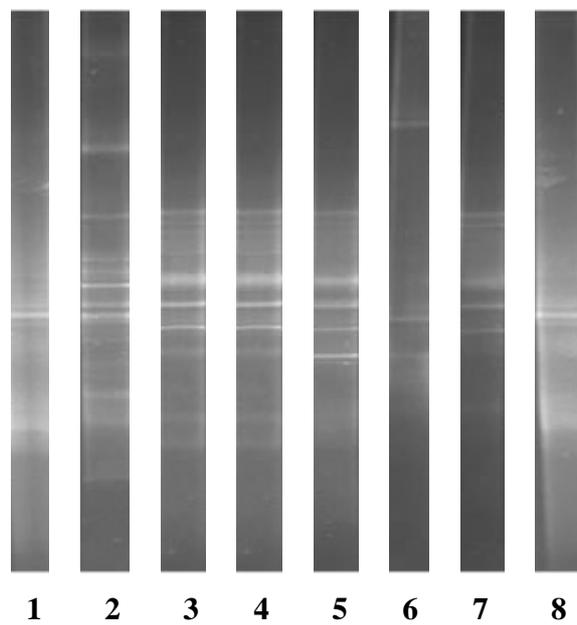


Figura 58: Carriles extraídos de la Imagen DGGE, de la prueba 3.

Al igual que en la prueba 1 y 2 el detalle de esta implementación se encuentra en el Anexo F.

En el siguiente recuadro señalado como Tabla 11, muestra la comparación resultante sobre la detección automática de bandas en ambos programas. Además se agrega una columna la cual representa las bandas reales, las cuales fueron ratificadas por un biólogo

Carril	Bandas Detectadas con DIDB	Bandas Detectadas con Q1	Bandas Reales
1	4	12	4
2	3	11	4
3	5	12	5
4	8	12	8
5	7	8	7
6	10	14	10
7	2	8	3
8	2	9	2

Tabla 11: Cuadro comparativo de las Bandas encontradas en la Prueba 3.

Carril	Bandas Detectadas con DIDB		Bandas Detectadas con Q1		Bandas Reales
	Correctas	Erróneas y S/Detectar	Correctas	Erróneas y S/Detectar	
1	4	0	4	8	4
2	3	1	3	8 +1	4
3	5	0	5	7	5
4	8	0	5	7+3	8
5	7	0	2	6+5	7
6	10	0	7	5+3	10
7	2	1	3	5	3
8	2	0	2	7	2

Tabla 12: Cuadro comparativo con las Bandas Detectadas de manera exitosa y con error.

Carril	Probabilidad DIDB		Probabilidad Q1	
	Éxito	Error	Éxito	Error
1	1	0	0,33	0,67
2	0,75	0,25	0,18	0,82
3	1	0	0,417	0,583
4	1	0	0,167	0,833
5	1	0	0,083	0,917
6	1	0	0,429	0,571
7	0,67	0,33	0,375	0,625
8	1	0	0,222	0,778

Tabla 12: Cuadro comparativo con las probabilidades de éxito y de error en la detección de bandas en ambas aplicaciones

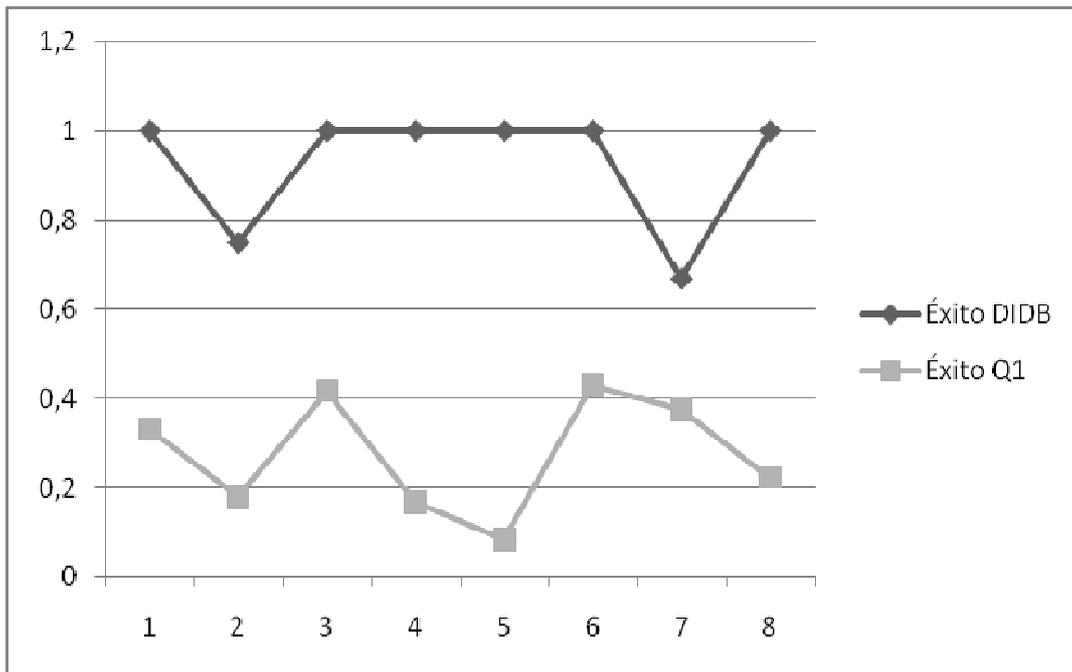


Figura 59: Gráfico comparativo entre el éxito en la detección en ambos programas.

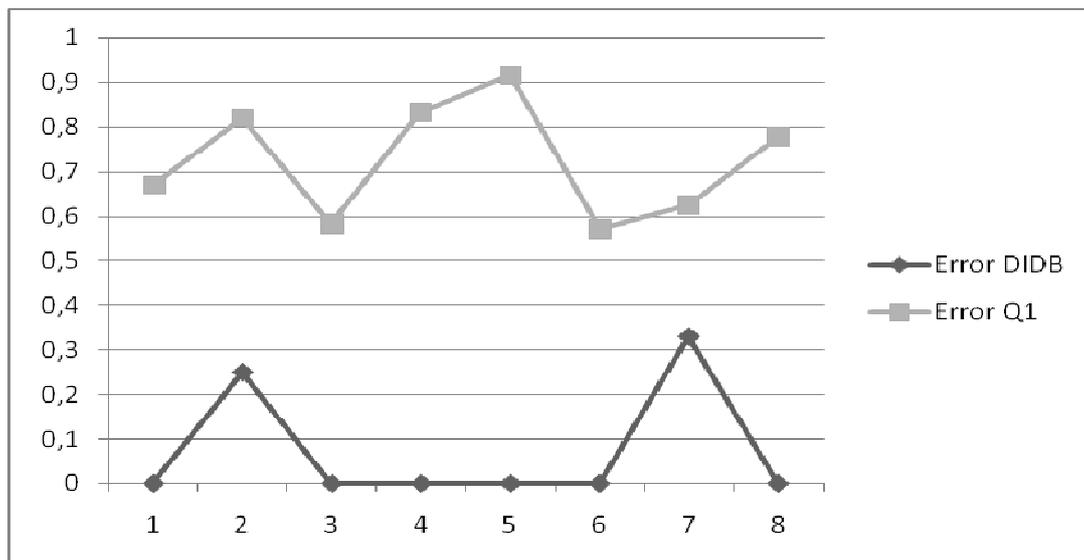


Figura 60: Gráfico comparativo entre el error en la detección en ambos programas.

Probabilidad de Éxito DIB $0,9275 \approx 92,75\%$

Probabilidad de Error DIB $0,0725 \approx 7,25\%$

Probabilidad de Éxito Q1 $0,275375 \approx 27,5375\%$

Probabilidad de Error Q1 $0,724625 \approx 72,4625\%$

En resumen:

	Éxito DIDB	Error DIDB	Éxito Q1	Error Q1
Prueba 1	0,91875	0,08125	0,370625	0,629375
Prueba 2	0,97325	0,02675	0,6889	0,31108125
Prueba 3	0,9275	0,0725	0,275375	0,724625
PROMEDIO	0,93983333	0,06016667	0,44496667	0,55502708

Tabla 13: Cuadro resumen con los resultados de las 3 pruebas .

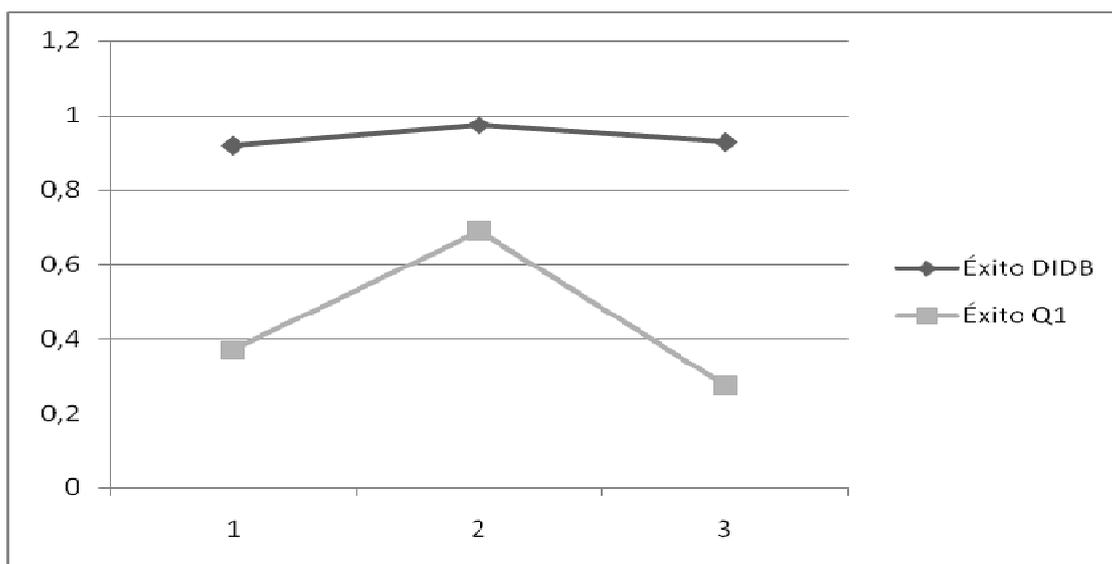


Figura 61: Gráfico comparativo sobre el éxito en la detección en las 3 pruebas.

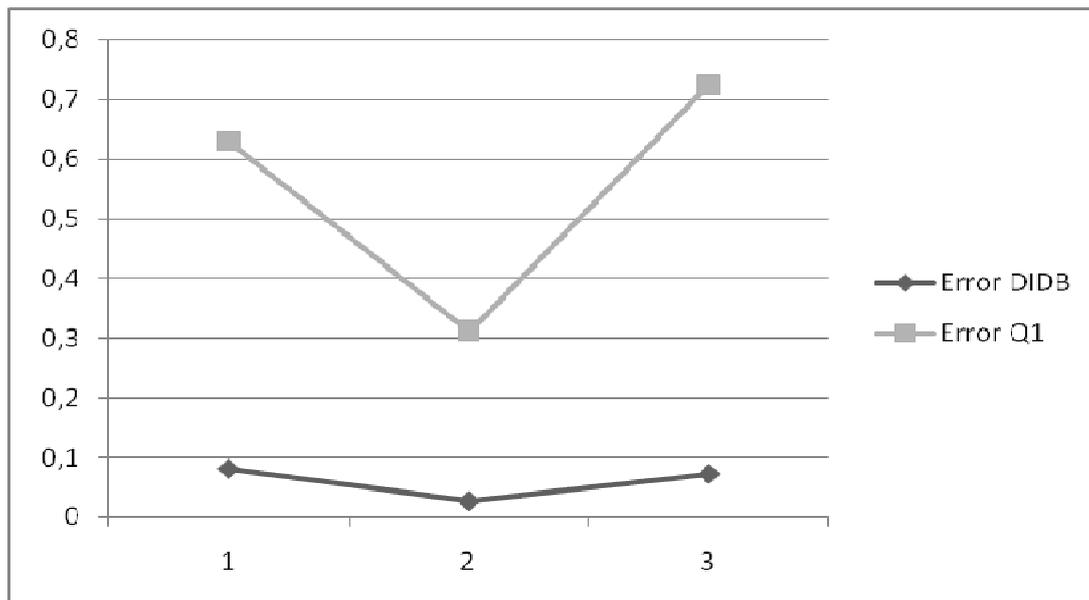


Figura 62: Gráfico comparativo sobre el error en la detección en las 3 pruebas.

Como se pudo observar en los gráficos representados por las Figuras 64 y 65, hay una gran diferencia con respecto a los resultados esperados sobre la detección de bandas al ejecutar ambas aplicaciones. Dejando de manifiesto con ello, que DIDB permite disminuir considerablemente la tasa de error del Q1 de un 55,5% a un 6,01%, además de aumentar la eficiencia en la detección de un 44,49% a un 93,98%. Lo que significa que la aplicación que se está desarrollando estaría cumpliendo a cabalidad lo esperado por los usuarios, sumándose a ello la garantía de poder procesar cualquier imagen en cualquier tipo o formato, como se ha explicado en las secciones anteriores de este capítulo.

4.6.2. Resultados del Módulo CDB

La manera en cómo se puede apreciar los resultados arrojados por el CDB, es a través de la ejecución de diversas pruebas en diferentes imágenes. Dando con ello los siguientes resultados esperados:

Cuantificación Prueba 1

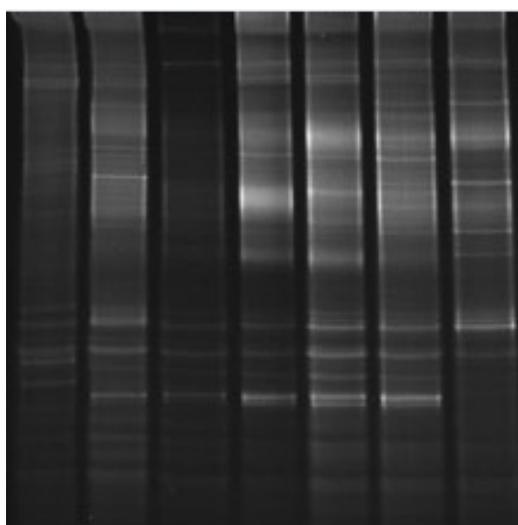


Figura 63: Imagen DGGE para la prueba 1.

Carril	Banda	Cuantificación	Participación en Carril
1	1	102,79	0,251099277
	2	102,19	0,249633574
	3	102,189	0,249631132
	4	102,191	0,249636017
	TOTAL	409,36	
2	1	109,89	0,196318733
	2	118,87	0,212361524
	3	114,78	0,205054729
	4	106,453	0,190178525
	5	109,76	0,196086488
	TOTAL	559,753	

3	1	102,79	0,200619091
	2	103,001	0,201030908
	3	102,189	0,199446097
	4	102,191	0,19945
	5	102,193	0,199453904
	TOTAL	512,364	
4	1	178,98	0,082817958
	2	199,0123	0,092087341
	3	176,98	0,081892514
	4	146,833	0,067942838
	5	189,9282	0,087883929
	6	389,98	0,180452269
	7	230,78	0,10678695
	8	220,89	0,102210631
	9	214,892	0,099435225
	10	212,85	0,098490347
	TOTAL	2161,1255	
5	1	198,32	0,095327822
	2	212,34	0,10206691
	3	235,876	0,113380119
	4	298,005	0,143244088
	5	304,892	0,146554509
	6	297,002	0,142761969
	7	189,003	0,090849356
	8	165,982	0,079783695
	9	178,98	0,086031532
	TOTAL	2080,4	
6	1	102,79	0,167201286
	2	102,19	0,166225308
	3	101,967	0,165862569
	4	102,191	0,166226934
	5	103,23	0,167917003
	6	102,4	0,1665669
	TOTAL	614,768	
	1	198,873	0,08575705
	2	201,982	0,087097698

7	3	189,472	0,081703196
	4	188,982	0,081491901
	5	301,918	0,130191615
	6	265,98	0,114694605
	7	278,674	0,12016845
	8	234,98	0,101326935
	9	245,987	0,10607332
	10	212,18	0,09149523
	TOTAL	2319,028	

Tabla 14: Participación y Cuantificación de las Bandas

Carril	Número de Bandas	Promedio	Participación en DGGE 1
1	4	0,04728769	0,08275346
2	5	0,06466051	0,09052471
3	5	0,05918631	0,08286084
4	10	0,24964489	0,17475142
5	9	0,24031979	0,18691539
6	6	0,07101563	0,08285157
7	10	0,26788518	0,18751962

Tabla 15: Cuadro Resumen de la Cuantificación de la Imagen DGGE 1.

Interpretaciones

En la imagen DGGE utilizada a modo de ejemplo, se puede apreciar que el número mayor de bandas, se encuentran dentro del carril 4, 5 Y 7, lo que conlleva a pensar que dentro de estos carriles existe mayor riqueza o diversidad de especies, si se les compara con el carril 1 en donde su riqueza es igual a 4, determinado por el número de bandas. Cabe señalar además que la participación de estos carriles, es mayor que las demás dentro de la imagen, ya que la presencia de bandas y su cuantificación sobrepasa a los demás carriles, si se realizará una sumatoria del total de ellas y se comparará entre ellos. Como se puede apreciar en la Tabla 15, en donde la participación mayor recae justamente en aquellos

carriles, obteniendo el siguiente porcentaje de participación de 17,47%, 18,69% y 18,75% para los carriles 4, 5 y 7 respectivamente.

Nota

El detalle de la segmentación de esta imagen se encuentra en el Anexo G.

4.6.3. Generación de Reportes

Otras de las implementaciones que se han realizado con respecto al sistema CYDB, es la generación de reportes, la cual fue desarrollada a partir de etiquetas HTML incrustadas dentro de un archivo en Matlab. Ya que, de esta manera se permiten realizar reportes más dinámicos, que los que se podrían esperar de uno en formato TXT o DAT, debido a que la utilización de estas etiquetas permiten incrustar imágenes DGGE, carriles y bandas, aparte de la información en texto que navega a través de variables globales de una función a otra.

Ofreciendo con ello la oportunidad de que al encontrarse con algún reporte de este tipo, el usuario pueda abrirlo a través de cualquier navegador, además de asociar los datos inmediatamente con la imagen DGGE insertada en él, evitando de esta manera que pierda tiempo en asociar el reporte con alguna imagen que se encuentre en otro lugar del equipo.

Un ejemplo de ello se puede apreciar en las siguientes líneas de código:

```
fwrite(arch, '<HTML>');
fwrite(arch, '<TITLE>');
fwrite(arch, '<H1>');
fwrite(arch, 'REPORTE IMAGEN DGGE 1');
fwrite(arch, '</H1>');
fwrite(arch, '</TITLE>');
fwrite(arch, '<BODY>');
fwrite(arch, '<H1>');
fwrite(arch, '<p align="center" >REPORTE IMAGEN DGGE 1</p>');
fwrite(arch, '</H1>');
fwrite(arch, '<H2>');
fwrite(arch, '<p align="center" >Autor: Carola Figueroa Flores</p>');
fwrite(arch, '</H2>');
fwrite(arch, '<H3>');
fwrite(arch, '<p align="center">ICINF UBB</p>');
fwrite(arch, '</H3>');
fwrite(arch, '<BR>');
fwrite(arch, '</BR>');
fwrite(arch, '<table width="100%" border="0" cellpadding="0" cellspacing="0">');
fwrite(arch, '<tr>');
```

```
fwrite(arch , '<th scope="col"></th>');
fwrite(arch , '<th scope="col">&nbsp;</th>');
fwrite(arch , '<th scope="col"></th>');
fwrite(arch , '</tr>');
fwrite(arch , '<TR>');
```

Líneas que demuestran las incrustaciones de etiquetas html a un archivo denominado en este caso como *arch*. Originando con ello lo siguiente, representado por la Figura 67.

REPORTE IMAGEN DGGE 1

Autor: Carola Figueroa Flores

ICINF UBB

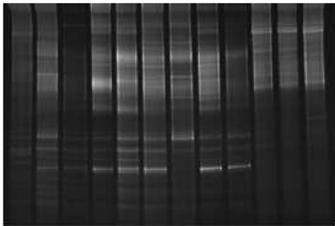
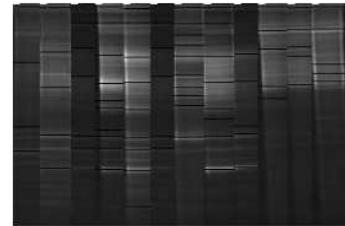


Imagen DGGE 1



Carriles Concatenados

Figura 64: Demostración de las etiquetas html dentro del lenguaje Matlab.

4.6.4. Desarrollo de Interfaz

Una cuarta implementación es el desarrollo de interfaz, la cual tiene que ver con la manera en que el usuario a futuro interactuará con el sistema CYDB, sin embargo la finalización de esta última implementación, no está dentro de los plazos fijados, es por ello que sólo ha sido posible diseñar el prototipo de interfaz representado en la Figura 65 y 66, dejando como tarea a futuro el poder finiquitarlas.

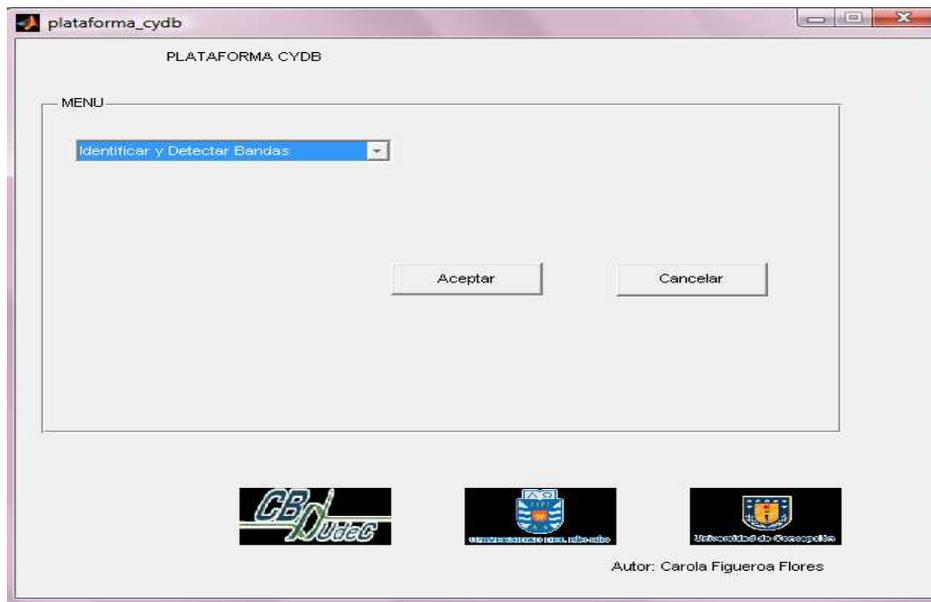


Figura 65: Plataforma del Sistema CYDB.

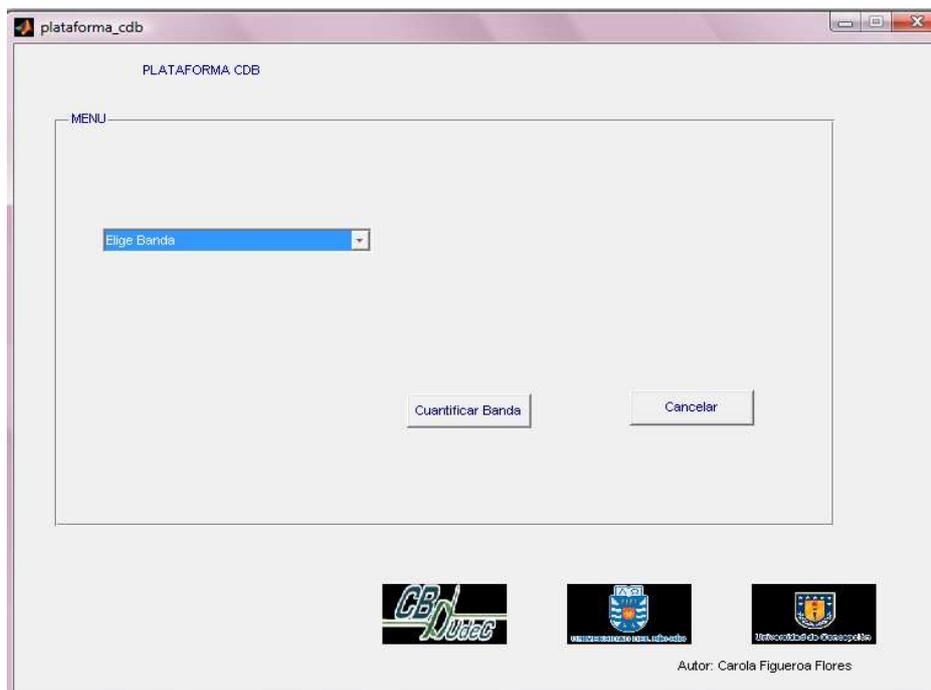


Figura 66: Plataforma del módulo CDB

Conclusiones

Luego de estudiar las posibilidades de desarrollar un sistema con las características necesarias, se puede dimensionar las potencialidades que ofrece el área de la Bioinformática. Siendo un área de la informática muy interesante de conocer, sobre todo para dar respuestas y soluciones en temas tan importantes de manejar, como es el caso del estudio del Genoma, Secuencias de ADN, Proteínas, etc. Es por esta razón que al ser un campo incipiente en las ciencias de la computación, abre una gama de posibilidades de desarrollo en nuevas investigaciones y aplicaciones, con el propósito de automatizar y agilizar procesos que en muchos centros de Biotecnologías aun se llevan manual o artesanalmente.

El procesamiento y Análisis de Imágenes DGGE, es un área que aún necesita de experimentación y seguimiento, con el fin de seguir afirmando o rechazando las teorías con respecto al número de diversidades de especies que existen en nuestro medio ambiente, y así seguir avanzando en el desarrollo de la genética.

Este proyecto en particular ha querido tomar como base los fundamentos de la ingeniería de software y complementarlos con los estudios realizados en el área de procesamiento de imágenes, con el fin de entregar un sistema más robusto, a partir de un adecuado análisis de requerimientos. Es por este motivo que se creó un capítulo completo que abordaba este tema, con el fin de avanzar a la siguiente etapa que consistía en el desarrollo del sistema desde el punto de vista de la implementación, y las metodologías involucradas en ello. Para ello se necesitó de realizar innumerables pruebas, las cuales no todas fueron descritas en este documento, quedando de manifiesto los exitosos resultados arrojados por el módulo de Detección e Identificación de bandas, en donde disminuía considerablemente la tasa de error arrojada por el Quantity-One, software utilizado hasta la fecha por el Centro de Biotecnología de la Universidad de Concepción. Siendo un aporte para el área de estudio, ya que además de disminuir el error, demostraba poseer características que el mismo Q1 no poseía, como

por ejemplo: permitir que cualquier imagen en cualquier formato o tipo, podría ser procesada, ya que este está restringido sólo a procesar imágenes TIF.

Cabe señalar que aparte de generar una automatización exitosa en la detección de estas Bandas, se logró entregar una propuesta de identificación, la cual además de entregar cifras, genera un reporte en formato HTML con una agradable interfaz para el usuario, el cual contendrá un detalle de la imagen recién procesada con la información de cada carril, además de imágenes que ayudan al usuario a percatarse a que DGGE se refiere cada reporte, ya que todo ello saldrá explícito en él.

Otro punto importante de este proyecto es que existen muchas posibilidades de seguir especializándose en este tema, ya que una de las principales causas por las cuales fue imposible agilizar la entrega de resultados de la cuantificación, se debió a la búsqueda de alguna técnica óptima, que permita segmentar cada banda del carril, sin que se tenga que perder información que podría ser relevante para los investigadores, y para ello se estuvo jugando con varias alternativas, siendo la más tentadora la concerniente al uso de la normalización de histogramas, ya que se podría disminuir considerablemente la entropía en decidir si lo que se veía en la imagen correspondía a background o a banda. Además de este hecho, es de gran importancia concretar la interfaz que se logró diseñar y mostrar en el documento, ya que es de vital importancia para el laboratorio que se terminará algún día.

Glosario

Esta sección fue creada con el fin de proveer al lector de una mejor comprensión sobre los términos empleados en este documento.

1. Electroforesis.

Método de laboratorio en el que se utiliza una corriente eléctrica controlada con la finalidad de separar Biomoléculas según su tamaño y carga eléctrica a través de una matriz gelatinosa.

2. Biopelículas.

Las biopelículas son organizaciones microbianas compuestas por microorganismos que se adhieren a las superficies gracias a la secreción de un exopolímero. Estas conformaciones microbianas presentan características como heterogeneidad, diversidad de microambientes, resistencia a antimicrobianos y capacidad de comunicación intercelular que las convierten en complejos difíciles de erradicar de los ambientes donde se establecen.

3. Microbiología.

La Microbiología, el estudio de los organismos microscópicos, deriva de 3 palabras griegas: mikros (pequeño), bios (vida) y logos (ciencia) que conjuntamente significan el estudio de la vida microscópica

4. Fouling.

Se refiere a la acumulación y deposición de material no deseado en superficies sólidas, con mayor frecuencia en un medio acuático. Esta acumulación se compone de cualquier organismo vivo o de compuestos orgánicos o inorgánicos.

5. Pixel.

Es el principal elemento de una imagen, por lo que la unión de todos ellos hace posible obtenerla.

6. Mapa de Bits.

Una imagen bitmap o mapa de bits, esta compuesta por pequeños puntos o pixeles con unos valores de color y luminancia propios. El conjunto de esos pixeles componen la imagen total.

7. Reacción en Cadena de la Polimerasa o PCR.

Técnica de la biología molecular, cuyo objetivo es obtener un gran número de copias de un fragmento de ADN en particular. Además permite amplificar un fragmento del ADN, pues de esta manera resulta más fácil o más sencillo reconocer como por ejemplo una bacteria un virus, entre muchas cosas más.

8. Tm.

Es un parámetro que permite determinar hasta que porcentaje, correspondiente al 50% de Adenina y Guanina se puede separar una secuencia de ADN, para no separarse completamente.

9. GCclamp.

Gclamp es una extensión de Citocina y Guanina que permite, que las hebras de las secuencias de ADN, no se separen completamente.

10. Primers.

Son los iniciadores de la técnica del PCR, también denominados cebadores, los cuales permiten aumentar el tamaño del ADN, ya que están compuestos por Oligonucleótidos que lo facilitan.

11. Peso Molecular.

El Peso Molecular o Pb permite a los investigadores o biólogos a determinar el tamaño exacto y fácil de la secuencia de ADN.

12. Blots.

Se refiere a un conjunto de técnicas que facilitan el análisis de moléculas biológicas.

13. Estructura Taxonómica.

Permite clasificar a los organismos a qué categoría o familia pertenecen, a partir de sus taxones.

14. Gradiente.

El gradiente formado por un agente desnaturante y con las demarcaciones previas correspondiente a la temperatura para la localización de los fragmentos dentro de cada uno de los carriles y situarlos en una determinada banda (al mismo nivel de la temperatura). Además de ser perpendicular a la dirección de la electroforesis.

15. Desnaturalización.

La desnaturalización de la molécula de ADN es la separación de las hebras complementarias por enlaces no covalentes, esta se puede lograr aplicando calor o denaturantes químicos y está influenciada por:

- 1.- Las uniones de hidrógeno formadas entre bases complementarias
- 2.- La atracción entre bases vecinas de la misma hebra o la complementaria.

16. Ruido en la Imagen.

Todas las imágenes tienen cierta cantidad de ruido, la cual se puede deber a la cámara o al medio de transmisión de la señal. Generalmente el ruido se manifiesta como píxeles aislados que toman un nivel de gris diferente al de sus vecinos.

Referencias Bibliográficas.

- [1]- Definición Bioinformática: <http://infobiochip.isciii.es/Textos/GuiaRec/glosario.htm>
- [2]- Conceptos de Bioinformática: http://zeus.dci.ubiobio.cl/~tguierr/archivos-bio/Bio_01.pdf
- [3]- Principales áreas de Bioinformática http://es.wikipedia.org/wiki/Procesamiento_digital_de_imagenes .
- [4]- Cuadro Comparativo: <http://www.elai.upm.es/spain/Investiga/GCII/personal/lsantos/L SANTOS.HTM#Matlab>
- [5]- Prueba1: <http://www.scq.ubc.ca/wp-content/uploads/2007/02/DGGE.jpg>
- [6]- Prueba 2: http://www.abdn.ac.uk/staffpages/uploads/bch052/aglover-dgge_20gel.jpg
- [7]- Bioinformática: <http://www.solociencia.com/biologia/bioinformatica-concepto.htm>.
- [8]- Área de Estudio: www.centrobiotecnologia.cl
- [9]- Desarrollo Computacional Imágenes: <http://www.etsimo.uniovi.es/vision/intro/node7.html>
- [10]- http://es.wikipedia.org/wiki/Procesamiento_digital_de_imagenes
- [11]- Aplicación Un Scan It: <http://www.silkscientific.com/usinfo.htm>
- [12]. Aplicación MCID: <http://www.mcid.co.uk/?gclid=CMiSpIDjnZQCFQJtFQodfzBztg>

ANEXOS

Anexo A

Descripción de la Institución y área de estudio.

El Centro de Biotecnología de la Universidad de Concepción se crea el día 16 de diciembre de 2002, para dar respuesta a la Estrategia de Desarrollo de la Región de los años 2000-2006, definida mediante la participación de la comunidad Regional. El edificio del Centro de Biotecnología fue inaugurado el 27 de septiembre de 2005, con la presencia de destacadas autoridades gubernamentales, académicas y empresariales. Este Centro de Biotecnología se encuentra ubicado, en el campus Concepción, en un edificio de 3.000 m², que cuenta con laboratorios equipados para la investigación, desarrollo e innovación en biotecnología y es producto del esfuerzo compartido entre la Universidad de Concepción y el Gobierno de la Región del Bío-Bío.

Misión.

Desarrollar investigación biotecnológica de excelencia.

Visión.

Ser una institución líder en investigación, formación, transferencia e innovación en biotecnología de recursos naturales renovables, para contribuir de manera eficaz, eficiente e innovadora, al desarrollo económico y social del país.

Objetivos Estratégicos.

- Contribuir a la formación de capital humano especializado en la aplicación de herramientas biotecnológicas.
- Realizar transferencia tecnológica al sector productivo.
- Generar redes de colaboración, a través de la interacción con grupos de investigación nacional e internacional.

- Promover y participar en el desarrollo y utilización de la Biotecnología como herramienta para el desarrollo económico y social del país.

Organigrama.

El organigrama que representa de mejor manera la organización correspondiente al centro de Biotecnología es la siguiente:

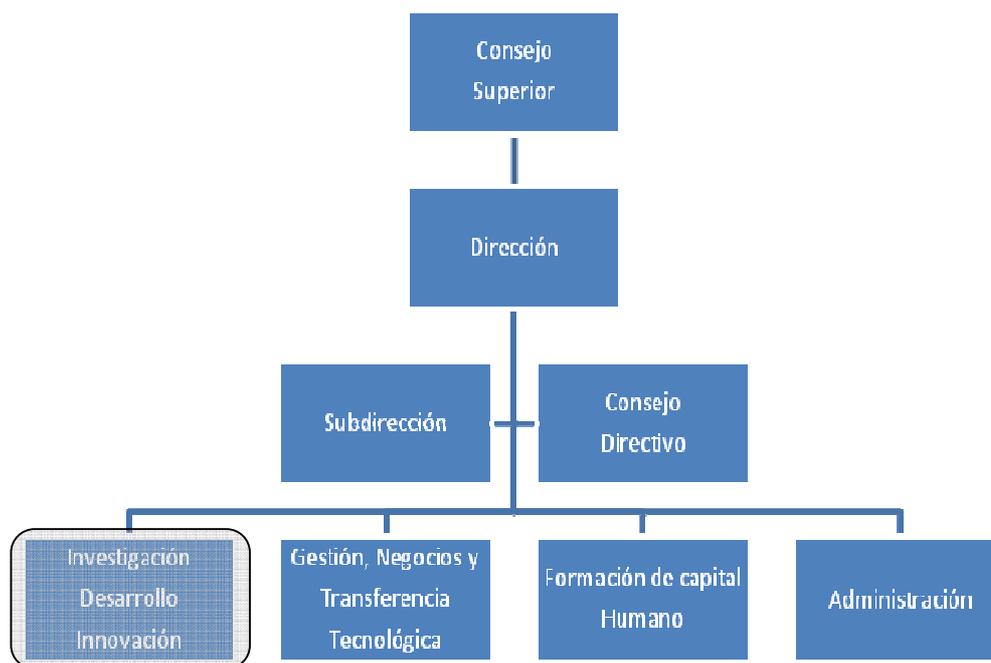


Figura 67: Organigrama del Centro de Biotecnología de la Universidad de Concepción.

En la base del organigrama, se pueden ver cuatro de las principales secciones del Centro de Biotecnología, siendo la de interés “Investigación Desarrollo Innovación”. Lugar en el cual se encuentra el área de estudio “El laboratorio de Biopelículas y Microbiología Ambiental”

Área de Estudio.

El área de estudio es el Laboratorio de Biopelículas y Microbiología Ambiental del Centro de Biotecnología de la Universidad de Concepción.

Descripción del área de estudio.

El laboratorio de Biopelículas y Microbiología Ambiental, cuenta con personal calificado y especializado, encargados de la investigación referente a la Biodiversidad de Microorganismos del ecosistema.

Funciones del Área de estudio

El laboratorio de Biopelículas y Microbiología Ambiental, se dedica principalmente a:

- Diseño y desarrollo de biopelículas para tratamiento biológico de efluentes industriales y de acuicultura.
- Biorremediación de suelos/sedimentos mediante barreras biológicas.
- Control de fouling y corrosión biogénica.
- Caracterización química y actividad biológica de bloqueadores transcripcionales en biopelículas de procariontes (Archaea E. Coli y Bacteria).
- Señalización celular y expresión de genes en biopelículas microbianas.
- Control adherencia microbiana y formación biopelículas de importancia agroforestal y acuícola.

Anexo B

Extracto del Documento: “Band Detection Random Amplified Polymorphism DNA Images” del Dr. Luis Ruedas

La copia original de este documento se encuentra en inglés, del cual

Once the image has been corrected and fit to a perfect rectangular template, the final task is removal procedure that eliminates the pixels which are in the quartile (25%) with lowest intensities. The first step consists of separating the lanes by creating a new sub-image for each lane, A_i , as the portion of the image A as follows: $A(x; y)$, where $x \in [b_{li}; b_{ri}]$, and y takes any value. Next step is to create a one-dimensional function with the horizontal running sum of pixel intensities as follows:

$$S_i(y) = \sum A_i(x, y)$$

The signal $S_i(y)$ is corrected by applying morphological operators (dilation, $S_i(y) \ominus q$, followed by erosion, $s_i(y) \oplus q$, with $q = [0; 1; 1; 1; 1; 1; 1; 0]$), followed by cubic spline smoothing correction. A function $s_{0i}(y)$ is obtained and the centers of the bands are located by finding the highest peaks (local maxima) of $s_{0i}(y)$. Here, we face two problems: (a) extremely low peaks that represent “false bands”, and (b) some high-intensity bands tend to “absorb” a band that is close to it. To correct (a), only those peaks above a threshold μ are considered, where μ is user-defined; we have experimentally found that $\mu = 200$ is a good value. To correct (b), the signal is normalized by taking a piecewise linear function as reference (baseline), where the knots of the normalizing function are selected as the lowest peaks of the signal. The signal on top represents the function $s_i(y)$, while the one in the bottom represents the smoothed signal (it has been plotted 1000 pixels below $s_i(y)$ to observe the difference between them). The normalized $s_i(y)$ function is plotted, along with the baseline ($y = 0$), and the threshold line ($y = \mu$, where $\mu = 200$ for the example).

Anexo C

Extracto del informe final del Proyecto de Título sobre: “ Análisis de Imágenes y Metodología de Desarrollo DGGE “

Análisis de Imágenes.

El Análisis de Imágenes , consiste en la extracción de grandes volúmenes de información a través de aplicaciones relacionadas con la Inteligencia Artificial y Computación Evolutiva, en base a algoritmos que permiten leerla una vez capturada en formato digital. La cual podría ser en escala de grises, o a color RGB (rojo, verde y azul), como se pueden apreciar en la Figura 68 y 69 respectivamente.



Figura 68: Imagen a escala de grises.



Figura 69: Imagen a color (RGB)

Desarrollo Computacional en el Análisis de Imágenes

En un principio cuando no existían las imágenes ni mucho menos las fotografías, la única herramienta posible de utilizar para memorizar alguna situación en particular, era a través de un simple bosquejo. Sin embargo, con la aparición de la fotografía, todo esto quedó de lado, permitiendo retener una copia fiel de un hecho real, a través de una imagen o fotografía, mejorando con ello la calidad de la información obtenida hasta entonces.

El desarrollo del uso de las técnicas de imágenes, se remonta a los años veinte del siglo XX, en el uso de la fotografía para el desarrollo de la información periodística entre Inglaterra y Estados Unidos. Dando origen a las primeras reproducciones de imágenes, por medio de las tonalidades

grises o monocromáticas. En las siguientes décadas el uso de esta técnica, se hacía a través de computadores, para la transmisión y procesamiento de imágenes. Lo que permitió a fines del siglo veinte, la incorporación de la Robótica, en el Análisis de Imágenes, consiguiendo una mejor automatización de los procesos involucrados en ello. Favoreciendo con ello, el desarrollo de los campos de la Geología, la Biología, la Criminología y la Medicina [7]

Imágenes digitales.

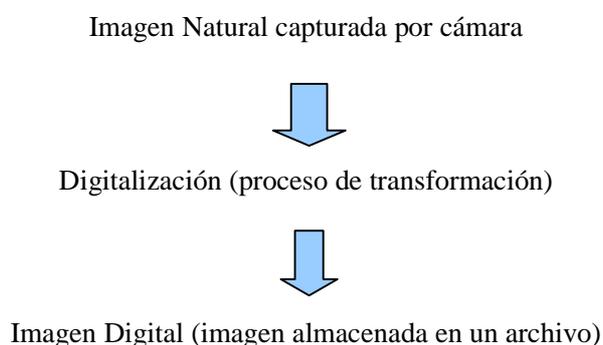
Una imagen digital puede ser una fotografía, un dibujo, un gráfico o cualquier tipo de figura, que se encuentre almacenada en algún computador, previamente efectuado el proceso de escaneamiento. Este tipo de imagen, puede estar representada en escala de grises o a color, siendo la primera una matriz de dos dimensiones de un orden $M \times N$, cuyos valores van del 0 al 255 (negro y blanco), representando dicho número al umbral de intensidad de un punto o pixel. En cambio una imagen a color, está conformada por tres matrices de un orden de $M \times N$ cada una, con un rango numérico de intensidad del 0 al 255, a diferencia de la primera, ésta permite la combinación de las 3 gamas de color (roja, azul y verde) [8].

Tipos de Procesos utilizados en una imagen digital.

En una imagen digital se pueden realizar los siguientes tipos de procesos:

- **Digitalización de la Imagen.**

Es el primer paso a realizar con el fin de lograr una imagen digital, y poder almacenarla en el disco duro del computador, o cualquier otro medio de almacenamiento, proceso por el cual puede observarse en el siguiente esquema.



- **Procesamiento de la Imagen.**

Una vez obtenida la imagen digital, es posible aplicar diferentes tipos de procesos con el fin de realzar sus colores o tonalidades, ajustar el brillo o contraste y por último disminuir el ruido provocado por la luz del flash.

- **Análisis de Imágenes.**

En este tipo de análisis es permitido realizar las siguientes funcionalidades, como por ejemplo: Extracción de características de la imagen, Interpretaciones múltiples, Reconocimiento de figuras, entre otras.

3.1.2.2. Elementos de una imagen digital.

La imagen digital posee los siguientes elementos que son propios de su caracterización.

- Tamaño de la Imagen, determinada por el número de píxeles.
- Intensidad de los niveles de gris.
- Tamaño del archivo, determinado en número de bytes.
- Tipo o formato del archivo, el cual podría ser en JPEG, GIF, PNG.

Clasificación de imágenes digitales.

La clasificación de las imágenes digitales se divide en dos grandes grupos, los cuales se diferencian en su tipo de formato y en cuanto al tratamiento que se les dé durante su análisis.

1. *Imagen Vectorial*, este tipo de imagen representa a las distintas formas geométricas, como el círculo, cuadrado, triángulo, etc., como se puede ver en la Figura 70. Permite con ello la utilización de operaciones Matemáticas, como la ampliación, reducción o expansión. Las cuales son muy utilizadas en los programas de dibujo, como Freehand, Corel Draw.
2. *Imagen en Mapa de Bits*, conforman un conjunto de puntos denominados píxeles, los cuales se encuentran posicionados cada uno en una matriz y de forma individual van describiendo la intensidad de un color en específico, como se puede observar en la Figura 74, la cual representa el Mapa de Bits del círculo. Comúnmente es utilizado este tipo de imagen cuando se quiere representar una imagen más parecida a la realidad o a un objeto en particular.

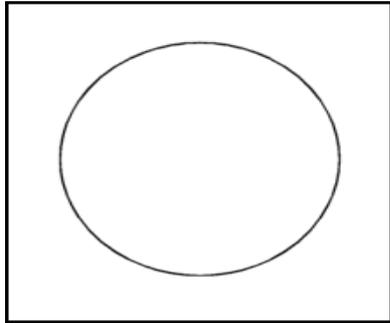


Figura 70: Imagen Vectorial de un círculo

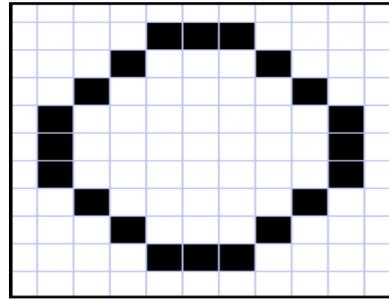


Figura 71: Imagen en Mapa de Bit de un círculo

Utilización de Histogramas.

El *histograma* de una imagen consiste generalmente en un gráfico de barras, donde se muestra el número de píxeles de cada nivel de gris que aparecen en la imagen. En donde las abscisas (eje X) representan a los distintos colores de la imagen y las ordenadas (eje Y) la frecuencia relativa a un punto las abscisas

El histograma proporciona información sobre el brillo y el contraste de la imagen, y puede ser utilizado para ajustar estos parámetros, eliminando así ciertas tonalidades molestas causadas por algún ruido de la imagen. En la Figura 72 se puede representar de mejor manera lo antes mencionado [9].

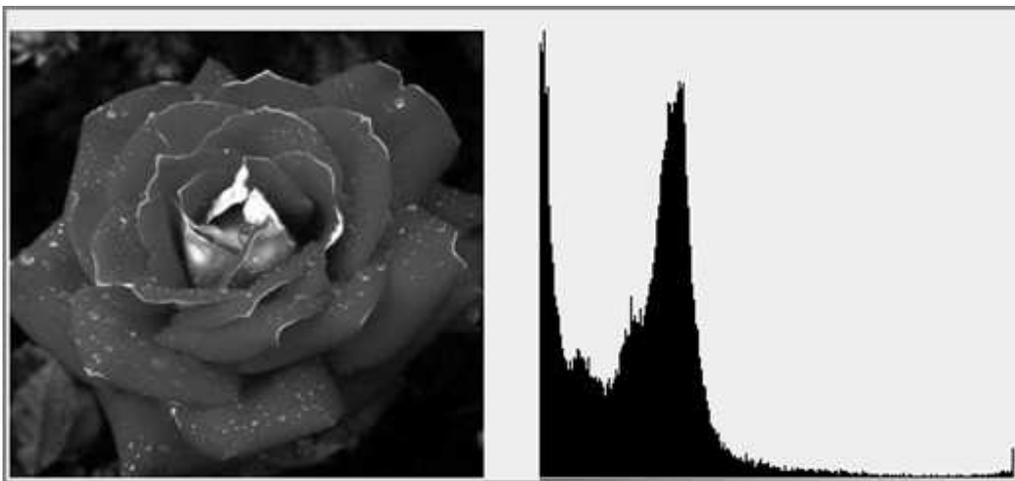


Figura 72: Ejemplo de Histogramas de frecuencia de intensidades en escala de grises

Imágenes DGGE.

DGGE es la sigla de Denaturing Gradient Gel Electrophoresis o en español Gel de Electroforesis en Gradiente Desnaturalizante. Siendo una técnica muy utilizada por el Ceb-UDEC, para el análisis de comunidades a partir del ADN o ARN. Con el propósito de estudiar las diversas especies que existen en el ecosistema y poder realizar la comparación entre diferentes hábitats, es decir, determinar cuántas especies existen en un río y en un pantano, para después concluir cuáles de ellos es más diverso, en base a la cantidad de organismos encontrado[10]

Ventajas:

- Alta tasa de detección gracias a la sensibilidad que posee esta técnica.
- La metodología a utilizar es simple y no necesita de radioactividad.
- Los fragmentos de PCR, pueden ser aislados desde el gel y usados en secuenciación.

Desventajas:

- Depende demasiado de análisis computacionales y de experimentos preliminares.
- Requiere equipamiento especial, para su desarrollo.
- Los primeros son más caros por el GC clamp, además se pueden requerir más partidores o primers para la secuenciación.
- Análisis de productos de PCR sobre 400 pb son menos exitosos
- Genes muy ricos en Guanina y Citosina no son fáciles de analizar por DGGE.
- El método involucra el uso de formamida para la decantación de las secuencias.

Tipos de tecnologías involucradas

Para este tipo de técnica, han surgido diversas alternativas tecnológicas, siendo en su mayoría aplicaciones que se comercializan en el exterior, debido al potente desarrollo en la Bioinformática en países tales como: Estados Unidos, España, Alemania y Bélgica entre otros. Siendo algunas de ellas las que se describirán a continuación:

1. UN- SCAN-IT Gel .

El SCAN -IT Gel es un software que sirve como apoyo en el análisis de imágenes de cualquier tipo de gel, una vez que ésta haya sido escaneada y transformada a formato digital [7].

i). Características de SCAN -IT Gel

- Digitaliza automáticamente los carriles y los segmentos de electroforesis en geles.
- Mejora la precisión y reproducibilidad de la electroforesis en gel de análisis.
- Elimina la necesidad de conjeturas o supuestos
- Reduce los tiempos de gel de análisis de laboratorio y aumentar la productividad.
- Determinar la abundancia relativa y la posición de cada grupo o segmento en el gel.
- Calibración de la imagen la intensidad de normalizar los resultados.
- Guardar los datos en formato ASCII, y en la exportación de hoja de cálculo, análisis de datos y programas gráficos.
- Digitalizar (x, y) con todos los gráficos de las mismas características que regular ONU-SCAN-IT.
- El gráfico y analizar (x, y) de datos con todas las características regulares de ONU-SCAN-IT

ii). Requisitos de Hardware:

Para la utilización de este tipo de software, los requisitos mínimos que debe poseer el computador, depende de la plataforma o del tipo de sistema operativo.

Plataforma WINDOWS

- PC compatible con IBM.
- Windows 95, 98, NT, 2000, Me, XP o Vista.
- 64 MB o más de memoria RAM.
- 30 MB de espacio libre en disco duro.
- Escáner, cámara digital, o de otro tipo de imagen comparable dispositivo.

Plataforma MAC

- Ordenador Macintosh compatible (Intel o basadas en PowerPC)
- Macintosh OS X 10,1 o superior.
- 64 MB o más de memoria RAM.
- 50 MB de espacio libre en disco duro.
- Escáner, cámara digital, o de otro tipo de imagen comparable dispositivo

iii). Valor de la Licencia.

El valor aproximado de la adquisición de la licencia es el siguiente:

- Para Sistemas Operativos Windows US \$445
- Para Sistemas Operativos MAC US \$445.
-

2. MCID : Gel & Blot Analysis System

El MCID es una poderosa herramienta de trabajo, que permite capturar y analizar, diferentes tipos de geles en formato digital [11]

i). Características de MCID.

- Permite adquirir y anotar: archivos e impresiones.
- Cuantifica imágenes de geles y blots.

ii) Requisitos de Hardware.

Los requisitos mínimos que debe poseer el computador para la instalación de este tipo de aplicación son los siguientes:

- Para plataformas Windows ® 2000 o Windows ® XP Pentium ® IV.
- Poseer un mínimo de 512 MB de RAM y 100 MB de disco duro.
- 128MB de tarjeta gráfica AGP.
- Pantalla SINGLE funcionando a 1024x768 píxeles.

- Para un rendimiento óptimo se recomienda Intel Procesador Core Duo, 2GB RAM, NVIDIA 256MB PCIe Graphics & 2 x 20 "Monitores LCD funcionando a 1600x1200 píxeles cada uno

3. Quantity –One.

Quantity One es un entorno de trabajo, que provee de variadas funcionalidades al usuario, para el análisis y procesamiento de imágenes de gel de electroforesis, blots[12].

i). Características del Quantity-One.

- El software se ejecuta en un equipo Windows o Macintosh.
- Utiliza una interfaz gráfica estándar con menús desplegables, barras de herramientas, y comandos de teclado.
- Permite cuantificar y analizar una cantidad de datos biológicos, entre ellos radiactivos, fluorescentes, y muestras adquiridas de los densitómetros, fósforo imágenes, fluorescentes de imágenes, gel y sistemas de documentación.
- La licencia gratuita tiene una duración de 30 días, y se puede obtener una vez registrado en el sitio de Bio Rad.

Anexo D

Funciones de Image Processing Toolbox de Matlab

Nombre	Sintaxis	Descripción
IMREAD	<pre>fileName = 'carriles/UBC 237' A = imread(fileName);</pre>	<p>A: es la variable que guarda a la imagen desde archivo.</p> <p>FileName: El nombre del archivo, que contiene al primer carril de la imagen UBC 237</p>
RGB2GRAY	<pre>A = imread('imagen.jpg'); G = rgb2gray(A);</pre>	<p>A almacena la imagen original, que supuestamente está a color.</p> <p>G almacena la imagen A en escala de grises gracias a rgb2gray.</p>
IMWRITE	<pre>imwrite(A, strcat(fileName, 'carril_1.jpg'), 'JPEG');</pre>	<p>A es la imagen que se guardará en el archivo fileName, con el nombre de Carril_1.jpg.</p> <p>Y strcat permite concatenar la imagen A con la extensión JPEG.</p>
IMSHOW	<pre>imshow(A');</pre>	<p>A es la imagen que se mostrará por pantalla.</p>
SIZE	<pre>[nx ny] = size(A');</pre>	<p>A es la imagen que se quiere saber sus dimensiones.</p> <p>[nx ny] es el arreglo que contiene el tamaño de las coordenadas x e y.</p>
IMROTATE	<pre>angle = 90°; R = imrotate(A, angle);</pre>	<p>R guardará la rotación de la imagen A, en un ángulo en este caso de 90°.</p>

Anexo E

Funciones de Statics Processing Toolbox

Nombre	Sintaxis	Descripción
PDF	<code>p=pdf('Normal',bandS,media,desv_standar);</code>	p es la variable que almacenará el área calculado por PDF. BandS = Contiene a la curva de una de las Bandas Normalizadas. Y los demás parámetros son los estimadores estadísticos calculados.
NORMCDF	<code>D=normcdf(bandS,media,varianza);</code>	D: es la variable que almacenará el área calculado por PDF. BandS = Contiene a la curva de una de las Bandas Normalizadas. Y los demas parámetros son los estimadores estadísticos calculados.
VAR	<code>varianza = var(sum(BANDA));</code>	Varianza es la variable en donde se guardará el cálculo de la varianza. Sum(BANDA): Son todos los valores de intensidades de cada Banda.
MEAN	<code>Media = mean(sum(BANDA));</code>	Media: es la variable en donde se guardará el cálculo de la media. sum(BANDA): Son todos los valores de intensidades de cada Banda.

Anexo F

Detalle de los resultados arrojados por el Módulo DIBB

IMPLEMENTACION PRUEBA 1

Carril 2



Uso del Módulo DIBB



- Uso del Quantity-One



Carril 3



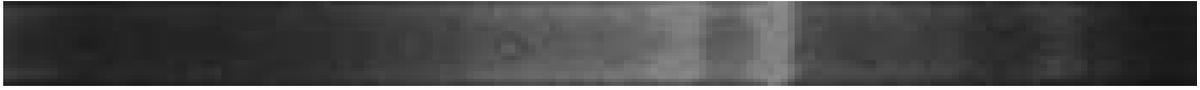
- Uso del Módulo DIBB



- Uso del Quantity-One



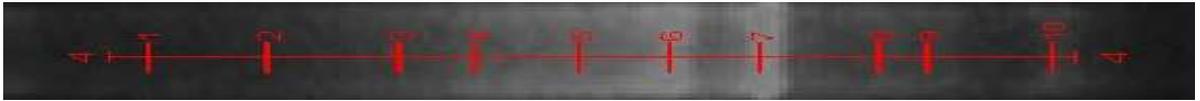
Carril 4



- Uso del Módulo DIDB



- Uso del Quantity-One



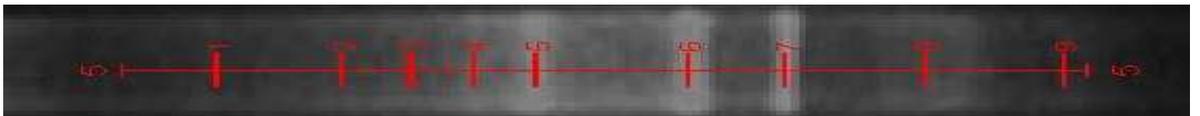
Carril 5



- Uso del Módulo DIDB



- Uso del Quantity-One



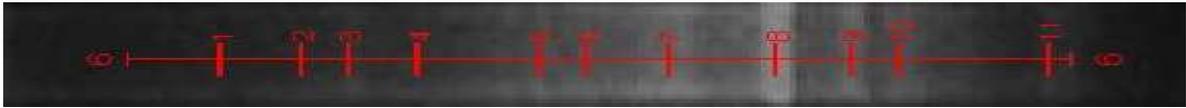
Carril 6



- Uso del Módulo DIDB



- Uso del Quantity-One



Carril 7



- Uso del Módulo DIDB



- Uso del Quantity-One

Bandas Identificadas: 8



Carril 8



- Uso del Módulo DIDB



- Uso del Quantity-One

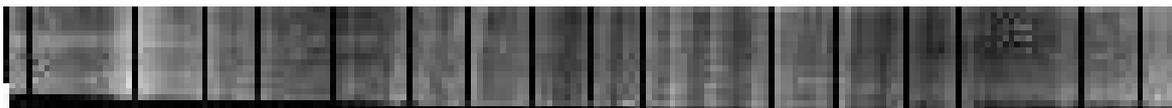


IMPLEMENTACION PRUEBA 2 DIDB

Carril 2



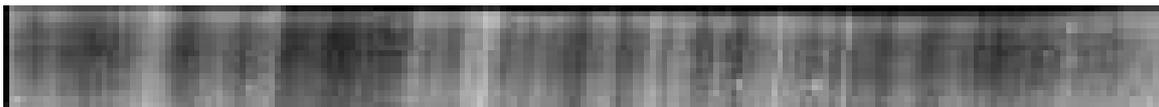
- Uso del Módulo DIDB



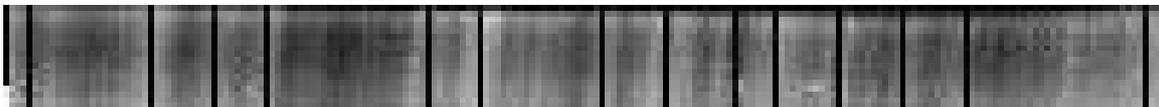
- Uso del Quantity-One



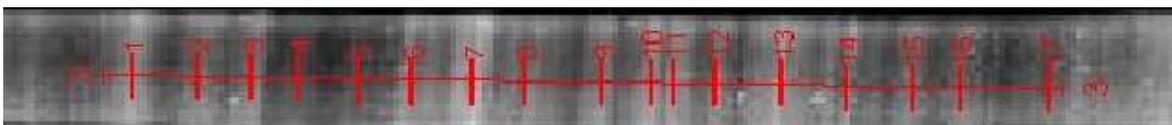
Carril 3



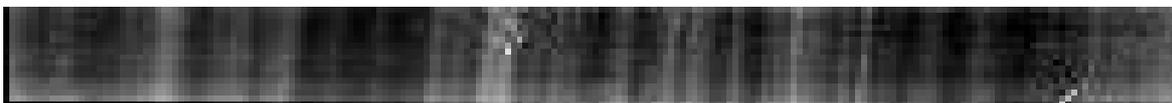
- Uso del Módulo DIDB



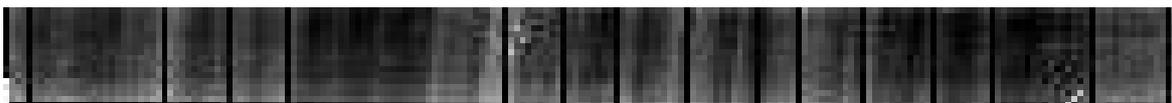
- Uso del Quantity-One



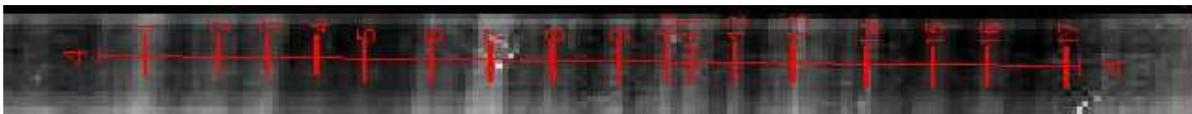
Carril 4



- Uso del Módulo DIDB



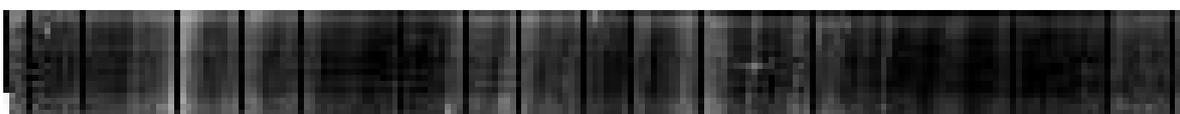
- Uso del Quantity-One



Carril 5



- Uso del Módulo DIDB



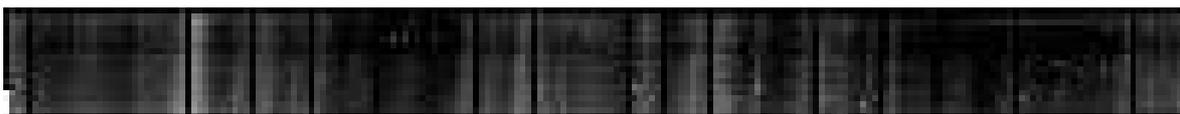
- Uso del Quantity-One



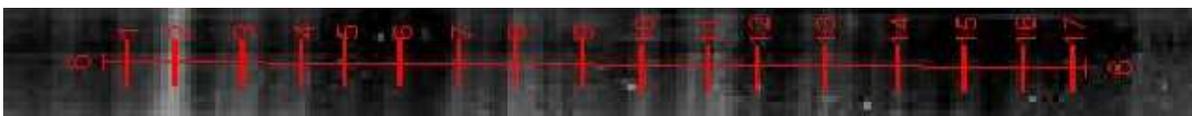
Carril 6



- Uso del Módulo DIDB



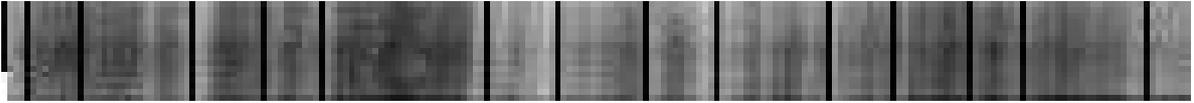
- Uso del Quantity-One



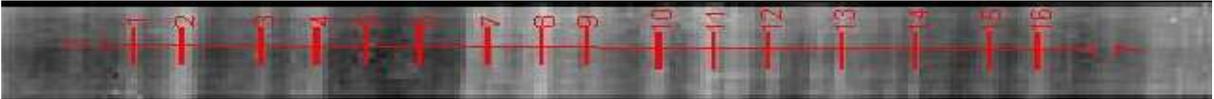
Carril 7



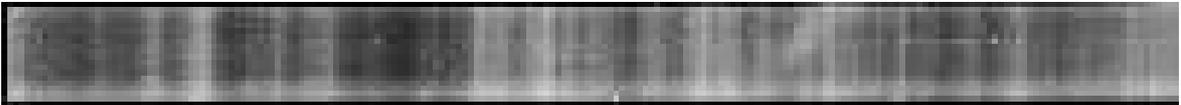
- Uso del Módulo DIDB



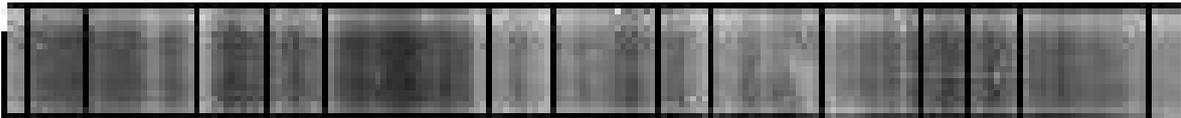
- Uso del Quantity-One



Carril 8



- Uso del Módulo DIDB



- Uso del Quantity-One



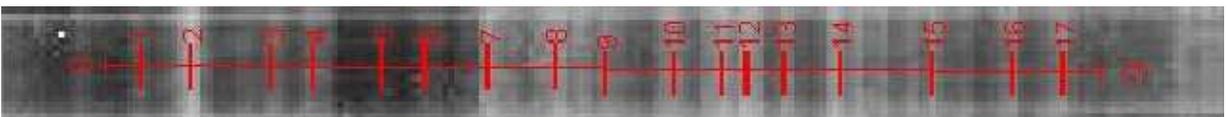
Carril 9



- Uso del Módulo DIDB



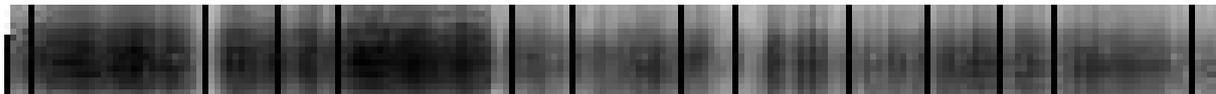
- Uso del Quantity-One



Carril 10



- Uso del Módulo DIDB



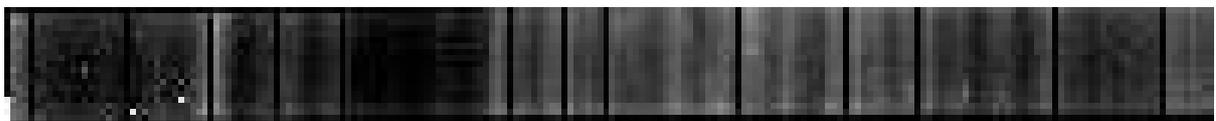
- Uso del Quantity-One



Carril 11



- Uso del Módulo DIDB



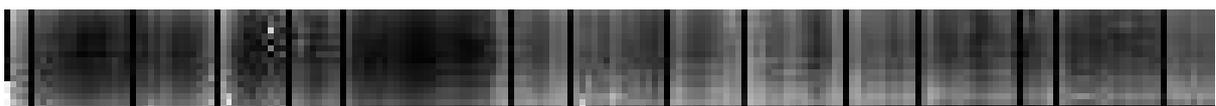
- Uso del Quantity-One



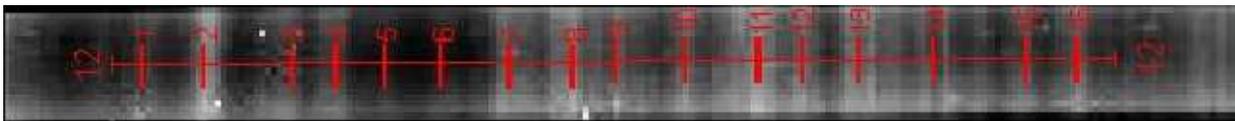
Carril 12



- Uso del Módulo DIDB



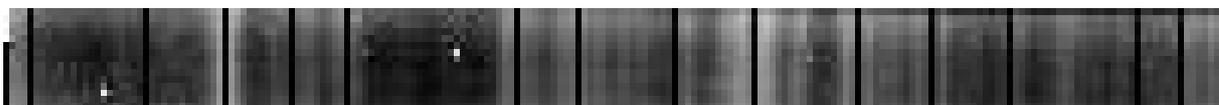
- Uso del Quantity-One



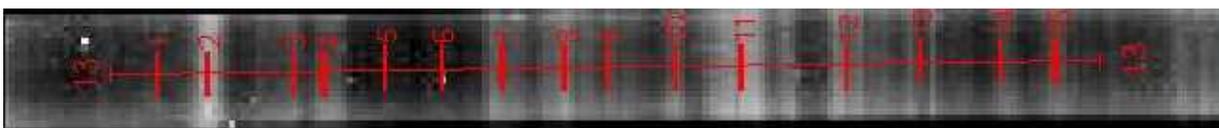
Carril 13



- Uso del Módulo DIDB



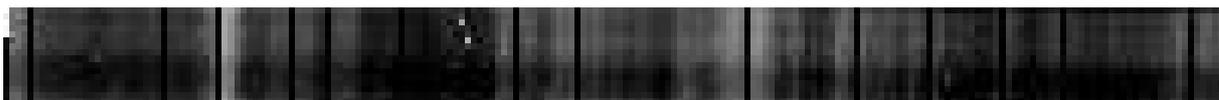
- Uso del Quantity-One



Carril 14



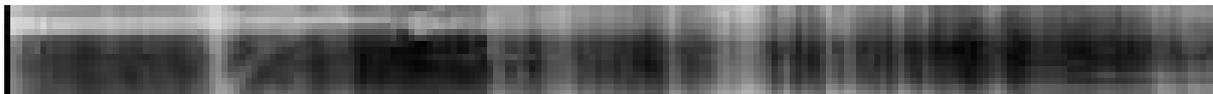
- Uso del Módulo DIDB



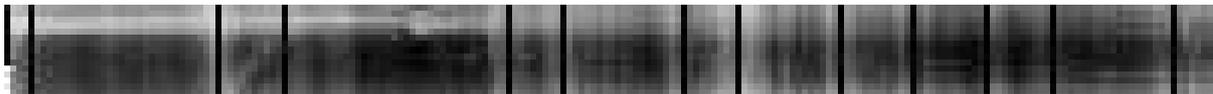
- Uso del Quantity-One



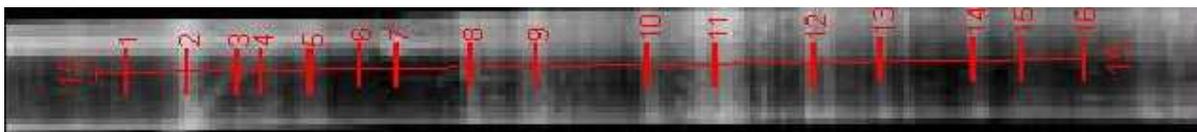
Carril 15



- Uso del Módulo DIDB



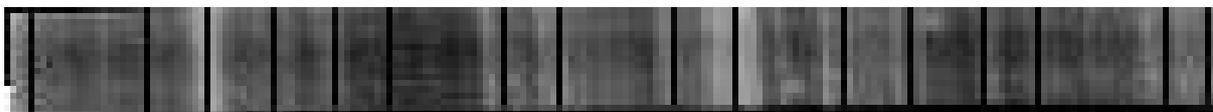
- Uso del Quantity-One



Carril 16

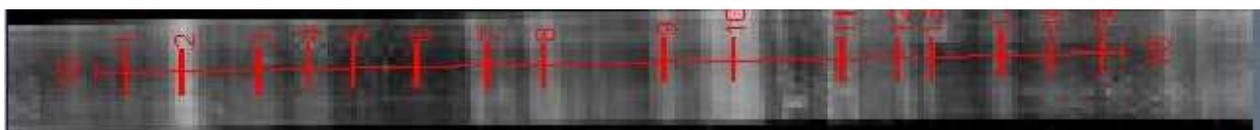


- Uso del Módulo DIDB



- Uso del Quantity-One

Bandas Identificadas: 16



IMPLEMENTACIÓN PRUEBA 3

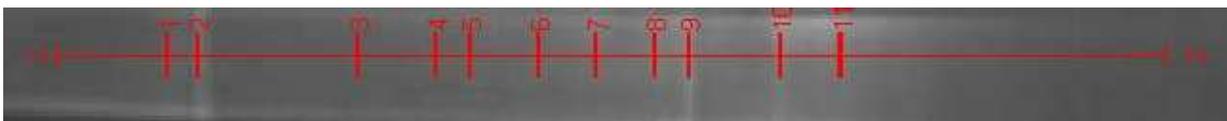
Carril 2



- Uso del Módulo DIDB



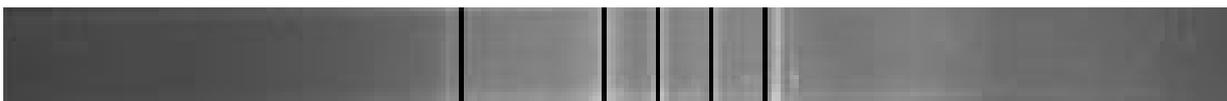
- Uso del Quantity-One



Carril 3



- Uso del Módulo DIDB



- Uso del Quantity-One



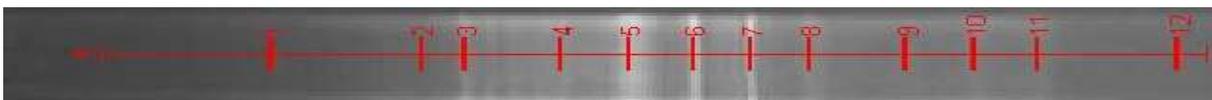
Carril 4



- Uso del Módulo DIDB



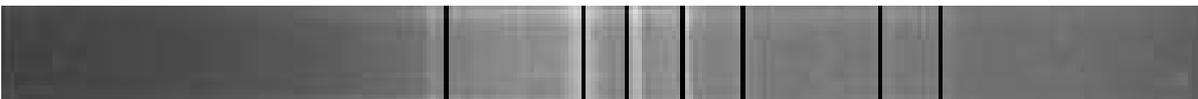
- Uso del Quantity-One



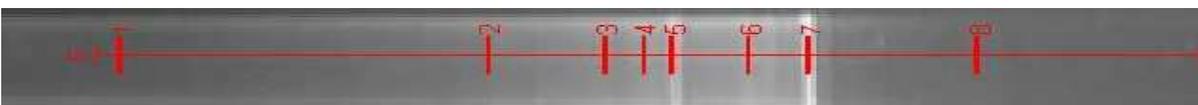
Carril 5



- Uso del Módulo DIDB



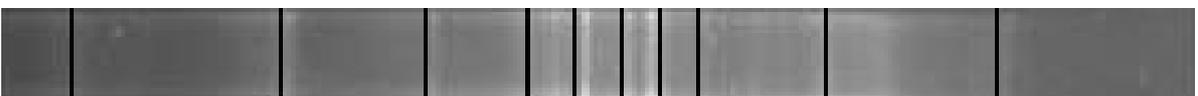
- Uso del Quantity-One



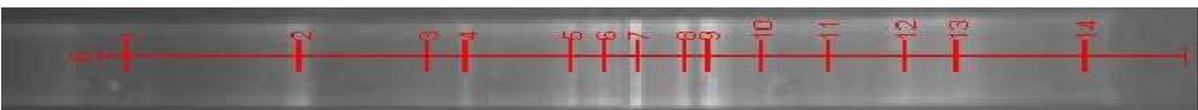
Carril 6



- Uso del Módulo DIDB



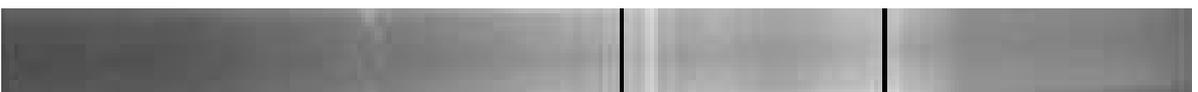
- Uso del Quantity-One



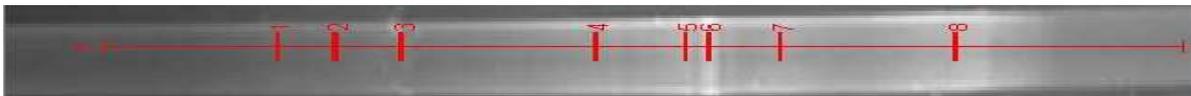
Carril 7



- Uso del Módulo DIDB



- Uso del Quantity-One



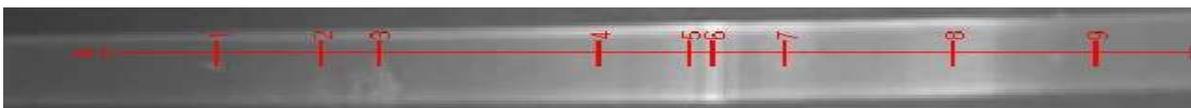
Carril 8



- Uso del Módulo DIDB



- Uso del Quantity-One



Anexo G

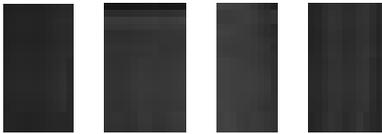
Aplicación del Módulo CDB

Segmentación de Bandas imagen Prueba 1

Carril 1



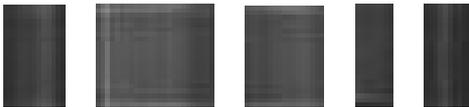
N° de Bandas: 4



Carril 2



N° de Bandas: 5



Carril 3



N° de Bandas: 5



Carril 4



N° de Bandas: 10



Carril 5



N° de Bandas: 9



Carril 6



N° de Bandas: 5



Carril 7



N° de Bandas: 10

